

OM of: US-09-303-518d-465 to: SPTRMBL_19:* out_format: pfs

Date: Jun 30, 2002 8:10 AM

About: Results were produced by the GenCode software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framed_ntp.model -DEV=xlh
-Q=/gen2.1/USPTO-spool/US09303518/runat_28062002.142714.4338/app-query.fasta.1.23501
-DB=SPTRMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELPO=6.000 -DELEXT=7.000 -START=1 -MATRIX=bls0sum62
-TRANS=human0.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09303518 -CGN1.1.1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-465

Query length: 1671

Database: SPTRMBL_19:

Database sequences: 562222

Database length: 172994929

Search time (sec): 883.700000

Score list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
SP_bacteriap:09K0G2	+ 2227.00	2717.34	8.5e-144	608	Q9K0G2 neisseria meningitidis
SP_bacteriap:09K120	+ 753.00	914.38	3.0e-43	467	Q9K120 neisseria meningitidis
SP_bacteriap:09K0V9	+ 753.00	886.18	1.0e-41	498	Q9K0V9 neisseria meningitidis
SP_bacteriap:09K0F4	+ 662.50	804.51	4.3e-37	422	Q9K0F4 neisseria meningitidis
SP_bacteriap:09K0P3	+ 454.00	547.27	7.6e-23	509	Q9K0P3 neisseria meningitidis
SP_bacteriap:09K0K8	+ 419.50	505.55	1.7e-20	482	Q9K0K8 neisseria meningitidis
SP_bacteriap:09K0X2	+ 474.50	487.15	1.8e-19	483	Q9K0X2 neisseria meningitidis
SP_human:09K0U0	+ 173.50	197.47	0.0012	956	Q9K0U0 homo sapiens (human)
SP_human:09K0U8	+ 173.50	188.86	0.0016	926	Q9K0U8 homo sapiens (human)
SP_human:09K0U5	+ 173.50	187.08	0.0016	926	Q9K0U5 homo sapiens (human)
SP_fungi:09P6T1	+ 170.00	185.97	0.0026	1992	Q9P6T1 neurospora crassa
SP_invertebrate:09N4S7	+ 167.50	188.94	0.0033	1079	Q9N4S7 caenorhabditis elegans
SP_human:09K952	+ 159.50	167.37	0.0157	3570	Q9K952 homo sapiens (human)
SP_rhodent:09P070	+ 156.50	168.52	0.0222	2187	P70670 mus musculus (mouse)
SP_bacteriap:09K0V9	+ 150.50	175.12	0.0393	529	Q9K0V9 microbacterium ammoniaphilum
SP_human:01A4879	+ 149.50	173.64	0.0463	543	Q1A4879 homo sapiens (human)
SP_invertebrate:02Q0908	+ 149.00	169.59	0.0549	770	Q2Q0908 caenorhabditis elegans
SP_invertebrate:09S527	+ 147.00	163.50	0.0828	1116	Q9S527 caenorhabditis elegans
SP_invertebrate:09N3Y8	+ 145.50	154.36	0.1273	2344	Q9N3Y8 caenorhabditis elegans
SP_human:09N859	+ 145.00	159.46	0.1183	1312	Q9N859 homo sapiens (human)
SP_human:09H7N4	+ 144.50	168.08	0.1002	513	Q9H7N4 homo sapiens (human)
SP_human:01A4881	+ 144.00	165.57	0.1139	622	Q1A4881 homo sapiens (human)
SP_fungi:09G6V6	+ 143.00	146.12	0.2163	3571	Q9G6V6 schizosaccharomyces pombe
SP_rhodent:070511	+ 142.00	148.97	0.2271	2622	Q70511 rattus norvegicus (rat)
SP_bacteriap:053478	+ 141.00	164.93	0.1683	456	Q53478 neisseria gonorrhoeae
SP_fungi:078953	+ 140.50	157.43	0.2187	920	P78953 schizosaccharomyces pombe
SP_bacteriap:069385	+ 140.00	152.76	0.2635	1390	Q69385 streptococcus mutans
SP_bacteriap:069391	+ 140.00	152.31	0.2667	1455	Q69391 streptococcus mutans
SP_bacteriap:069397	+ 140.00	152.31	0.2667	1455	Q69397 streptococcus mutans
SP_invertebrate:09Y0M6	+ 140.00	149.26	0.2891	1984	Q9Y0M6 drosophila melanogaster
SP_rhodent:070495	+ 139.50	156.46	0.2542	897	Q70495 mus musculus (mouse)
SP_human:09Q0U36	+ 139.00	152.39	0.3015	1275	Q9Q0U36 homo sapiens (human)
SP_bacteriap:069382	+ 139.00	151.09	0.3120	1455	Q69382 streptococcus mutans
SP_human:015038	+ 139.00	149.09	0.3290	1783	Q15038 homo sapiens (human)
SP_human:060382	+ 139.00	149.05	0.3294	1791	Q60382 homo sapiens (human)
SP_invertebrate:09VPG1	+ 139.00	146.65	0.3510	2284	Q9VPG1 drosophila melanogaster
SP_virus:09Q0P86	+ 138.00	155.14	0.3174	851	Q9Q0P86 human herpesvirus 4 (lept)

seq_name: sp_bacteriap:09K0G2	seq_documentation_block:	PRT:	608 AA.
SP_fungi:094497	+ 137.50	157.72	0.3154
SP_invertebrate:09V515	+ 137.00	155.21	0.3589
SP_rhodent:063625	+ 137.00	148.52	0.4286
SP_virus:056854	+ 136.50	153.24	0.4024
SP_invertebrate:017921	+ 136.00	153.68	0.4232
SP_human:060585	+ 136.00	153.05	0.4304
SP_invertebrate:09Y0Y8	+ 136.00	147.65	0.4967
SP_rhodent:088737	+ 135.50	137.00	0.7014
SP_human:09P1Y6	+ 135.00	144.93	0.6048
SP_fungi:007229	+ 134.00	150.64	0.5887
SP_human:09G2M2	+ 134.00	147.65	0.6373
SP_human:09G2V6	+ 134.00	147.22	0.6447
SP_human:09G2V5	+ 134.00	147.01	0.6493
SP_human:09G487	+ 134.00	146.97	0.6490
SP_human:09G487	+ 134.00	146.87	0.6507
SP_human:09G485	+ 134.00	146.85	0.6510
SP_human:09G485	+ 134.00	146.74	0.6528
SP_human:09G483	+ 134.00	146.41	0.6587
SP_invertebrate:076602	+ 134.00	146.26	0.6613
SP_human:09H482	+ 134.00	142.73	0.7264
SP_human:09H4D8	+ 134.00	141.28	0.7548
SP_human:09H481	+ 134.00	141.05	0.7594
SP_human:09Y0Y9	+ 134.00	141.04	0.7596
SP_invertebrate:076894	+ 133.50	144.70	0.7335
SP_invertebrate:09K0U5	+ 133.50	142.29	0.7821
SP_invertebrate:09K0U1	+ 133.50	141.80	0.7923
SP_invertebrate:09Y076	+ 133.00	138.82	0.8575
SP_invertebrate:09N8G4	+ 133.00	148.76	0.7009
SP_rhodent:063661	+ 133.00	141.62	0.8472
SP_human:09Y520	+ 133.00	137.66	0.9412
SP_human:01A888	+ 132.50	159.36	0.5629
SP_human:09N5M8	+ 132.50	144.24	0.8411
SP_invertebrate:046097	+ 132.00	144.65	0.9589
SP_human:09G5J4	+ 131.50	148.98	0.8400
SP_human:09G5J4	+ 131.50	148.96	0.8400
SP_human:09X0X9	+ 130.50	141.03	1.118
SP_invertebrate:09G0M6	+ 130.50	137.56	1.129
SP_plant:09Y0W8	+ 130.00	144.89	1.13
SP_plant:09Y0B2	+ 130.00	144.85	1.13
SP_human:09Y0B2	+ 130.00	144.33	1.15
SP_fungi:09H9Y9	+ 130.00	143.58	1.17
SP_bacteriap:085018	+ 130.00	142.62	1.20
SP_human:09K0B3	+ 130.00	142.12	1.21
SP_fungi:09P3J0	+ 129.50	134.39	1.49
SP_bacteriap:09K6J8	+ 129.50	153.89	0.9460
SP_plant:09STR2	+ 129.50	149.76	1.06
SP_plant:023054	+ 129.00	151.49	1.07
SP_fungi:09P652	+ 129.00	146.94	1.21
SP_virus:09P652	+ 128.50	143.10	1.43
SP_fungi:09Q223	+ 128.50	142.30	1.46
SP_invertebrate:09PVS5	+ 128.50	140.06	1.55
SP_plant:0949Y9	+ 128.00	151.57	1.21
SP_invertebrate:001699	+ 127.50	142.33	1.65
SP_human:09H306	+ 127.50	141.98	1.67
SP_human:09H307	+ 127.50	141.97	1.67
SP_human:01A4651	+ 127.50	141.12	1.70
SP_human:09K0W9	+ 127.50	138.76	1.81

seq_name: sp_bacteriap:09K0G2	seq_documentation_block:	PRT:	608 AA.
SP_fungi:094497	+ 137.50	157.72	0.3154
SP_invertebrate:09V515	+ 137.00	155.21	0.3589
SP_rhodent:063625	+ 137.00	148.52	0.4286
SP_virus:056854	+ 136.50	153.24	0.4024
SP_invertebrate:017921	+ 136.00	153.68	0.4232
SP_human:060585	+ 136.00	153.05	0.4304
SP_invertebrate:09Y0Y8	+ 136.00	147.65	0.4967
SP_rhodent:088737	+ 135.50	137.00	0.7014
SP_human:09P1Y6	+ 135.00	144.93	0.6048
SP_fungi:007229	+ 134.00	150.64	0.5887
SP_human:09G2M2	+ 134.00	147.65	0.6373
SP_human:09G2V6	+ 134.00	147.22	0.6447
SP_human:09G2V5	+ 134.00	147.01	0.6493
SP_human:09G487	+ 134.00	146.97	0.6490
SP_human:09G487	+ 134.00	146.87	0.6507
SP_human:09G485	+ 134.00	146.85	0.6510
SP_human:09G485	+ 134.00	146.74	0.6528
SP_human:09G483	+ 134.00	146.41	0.6587
SP_invertebrate:076602	+ 134.00	146.26	0.6613
SP_human:09H482	+ 134.00	142.73	0.7264
SP_human:09H4D8	+ 134.00	141.28	0.7548
SP_human:09H481	+ 134.00	141.05	0.7594
SP_human:09Y0Y9	+ 134.00	141.04	0.7596
SP_invertebrate:076894	+ 133.50	144.70	0.7335
SP_invertebrate:09K0U5	+ 133.50	142.29	0.7821
SP_invertebrate:09K0U1	+ 133.50	141.80	0.7923
SP_invertebrate:09Y076	+ 133.00	138.82	0.8575
SP_invertebrate:09N8G4	+ 133.00	148.76	0.7009
SP_rhodent:063661	+ 133.00	141.62	0.8472
SP_human:09Y520	+ 133.00	137.66	0.9412
SP_human:01A888	+ 132.50	159.36	0.5629
SP_human:09N5M8	+ 132.50	144.24	0.8411
SP_invertebrate:046097	+ 132.00	144.65	0.9589
SP_human:09G5J4	+ 131.50	148.98	0.8400
SP_human:09G5J4	+ 131.50	148.96	0.8400
SP_human:09X0X9	+ 130.50	141.03	1.118
SP_invertebrate:09G0M6	+ 130.50	137.56	1.129
SP_plant:09Y0W8	+ 130.00	144.89	1.13
SP_plant:09Y0B2	+ 130.00	144.85	1.13
SP_human:09Y0B2	+ 130.00	144.33	1.15
SP_fungi:09H9Y9	+ 130.00	143.58	1.17
SP_bacteriap:085018	+ 130.00	142.62	1.20
SP_human:09K0B3	+ 130.00	142.12	1.21
SP_fungi:09P3J0	+ 129.50	134.39	1.49
SP_bacteriap:09K6J8	+ 129.50	153.89	0.9460
SP_plant:09STR2	+ 129.50	149.76	1.06
SP_plant:023054	+ 129.00	151.49	1.07
SP_fungi:09P652	+ 129.00	146.94	1.21
SP_virus:09P652	+ 128.50	143.10	1.43
SP_fungi:09Q223	+ 128.50	142.30	1.46
SP_invertebrate:09PVS5	+ 128.50	140.06	1.55
SP_plant:0949Y9	+ 128.00	151.57	1.21
SP_invertebrate:001699	+ 127.50	142.33	1.65
SP_human:09H306	+ 127.50	141.98	1.67
SP_human:09H307	+ 127.50	141.97	1.67
SP_human:01A4651	+ 127.50	141.12	1.70
SP_human:09K0W9	+ 127.50	138.76	1.81

seq_name: sp_bacteriap:09K0G2	seq_documentation_block:	PRT:	608 AA.
SP_fungi:094497	+ 137.50	157.72	0.3154
SP_invertebrate:09V515	+ 137.00	155.21	0.3589
SP_rhodent:063625	+ 137.00	148.52	0.4286
SP_virus:056854	+ 136.50	153.24	0.4024
SP_invertebrate:017921	+ 136.00	153.68	0.4232
SP_human:060585	+ 136.00	153.05	0.4304
SP_invertebrate:09Y0Y8	+ 136.00	147.65	0.4967
SP_rhodent:088737	+ 135.50	137.00	0.7014
SP_human:09P1Y6	+ 135.00	144.93	0.6048
SP_fungi:007229	+ 134.00	150.64	0.5887
SP_human:09G2M2	+ 134.00	147.65	0.6373
SP_human:09G2V6	+ 134.00	147.22	0.6447
SP_human:09G2V5	+ 134.00	147.01	0.64

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gilm M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
 RA Cotton M.D., Utecherback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gall J., Scarlato V., Maignani V., Piza M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 MC58.";
 RL SCIENCE 287:1809-1815(2000).
 DR EMBL: AE002419; AAF62313.1; --
 DR TIGR: NMB0643; --
 KM Complete proteome.
 SQ SEQUENCE 608 AA; 66592 MW; AC6D4599E358BC5 CRC64;

alignment_scores:
 Quality: 2227.00 Length: 543
 Ratio: 4.708 Gaps: 5
 Percent Similarity: 87.109 Percent Identity: 80.295

alignment_block:
 US-09-303-518D-465 x Q9K0G2 ..

Align seg 1/1 to: Q9K0G2 from: 1 to: 608

1 TTGGCATTTCGCCGAATATCCCTTATCTGTCATCTGACAGTGTG 50
 :::
 1 MetGlyIleSerArgLysIleSerIleuIleuSerIleuAlaValCys 17
 51 CCTGCCGATGCATGCACACGCTCAGATTGGCAACGATCTTTTATCC 100
 :::
 17 sleuProMetHisAlaHisAlaSerAspLeuAlaAspSerPheIleA 34
 101 GGCAGGTTCTGCAGCCGTCAGATTGCAACCGGCAAGGAATACACCTA 150
 :::
 34 rglInValIleuAspArgIleHisPheGluProAspGlyLysTyrHisIleu 50
 151 TTGGCCACGAGGGGGAACCTTGCAGCGGCAAGGTCATATCGATTGGG 200
 :::
 51 PheGlySerArgGlyGluIleuAlaGluArgSerGlyHisIleGlyLeuG 67
 201 AAGCATTAACAGCCATCACTGAGGACCTGTCTTCACAGAGGGGCCA 250
 :::
 67 yLysIleGlnSerHisGlnLeuGlyAsnMetIleGlnGlnAlaIai 84
 251 TTAAGCAAAATATCGCTACATTTGCGCTTTCCGATCAGGGGCAAGAA 300
 :::
 84 IeLysGlyAsnIleGlyTyrIleValArgPheSerAspHisGlyHisGlu 100
 301 GTTCATTCCTCCCTTGCACACGATGCTCAGATTGCGATTGATGAGAC 350
 :::
 101 ValHisSerProPheAspAsnHisAlaSerHisSerAspSerAspGluAl 117
 351 CGGTAGTCCCGTTGACGATTACGCTTATCCGATTCATTTGGAGCGGAT 400
 :::
 117 aGlySerProValAspGlyPheSerIleuTyrArgIleHisTyrAspGlyT 134
 401 AGCAACACCATCCGCGGCGATATGAGGGGCAAGGGGCGGCGGTAT 450
 :::
 134 yrgIHisHisProAlaAspGlyTyrAspGlyProGlnGlyGlyGlyTyr 150
 451 CCGCGTCCCAAGGGCGGAGGATATATACAGCTACGACATTAAGGGGT 500
 :::
 151 ProAlaProLysGlyAlaArgAspIleTyrSerTyrAspIleLysGlyVal 167
 501 TGCCCAAAATATCGCTTCAACCTGACGCAACGCGACGACCGGACAC 550
 :::

167 lAlaGlnAsnIleArgLeuAsnLeuThrAspAsnArgSerThrGlyGlnA 184
 551 GGCTTTCGACCGCTTTCACGATACCGGATAGTATGCTACGCAAGAGATA 600
 :::
 184 rgluValAspArgPheHisAlaGlySerMetLeuThrGlnIleVal 200
 601 GGGAGGAGATTCAACGCGCCACCGATACAGCCCGGAGTGGACGATC 650
 :::
 201 GlyAspGlyPheLysArgLysThrArgTyrSerProGlnLeuAspArgSe 217
 651 GGGCAATGCCCGCGAAGCTTTACAGGCGATGACATATCTGCAAAACA 700
 :::
 217 rglAsnAlaIleAlaGluAlaPheAsnGlyThrAlaAspIleValLysAsnI 234
 701 TCATGGCGCGGAGAGAAATTGCGGCGGAGGATGCGGTGAGGAT 750
 :::
 234 IeIleGlyAlaIleAlaGlyGluIleValGlyAlaGlyAspAlaValGlnGly 250
 751 ATNAGCAAGGCTCAACATCTGTTATGCAAGGCTTGGCTGCTTTC 800
 :::
 251 IleSerGluGlySerAsnIleAlaValMetHisGlyLeuGlyLeuLeuSe 267
 801 CACCGAAACAGATGCGCGCATCAACGATTGGCAGATATGGCGCAAC 850
 :::
 267 rThrGluAsnLysMetAlaArgIleAsnAspLeuAlaAspMetAlaGlnL 284
 851 TCANAGCTATGCGCGAGCAGCATCCGATTTGGCAGTCCAAACCCC 900
 :::
 284 euLysAspTyrAlaIleAlaIleAlaIleArgAspTyrAlaValGlnAsnPro 300
 901 AATCGCGCAACAGGCTACAGCGCTCAGCATATCTTTACGCGAGTCT 950
 :::
 301 AsnAlaAlaGlnGlyIleGluAlaValSerAsnIlePheMetAlaAlaI 317
 951 CCGCGTCAAGGAGATTGAGCTGTTGCGGGAATACGCTTGGCGGCGCA 1000
 :::
 317 eProIleLysGlyIleGlyAlaValArgLysTyrGlyLeuGlyGlyI 334
 1001 TCAGGCGACATCTGTCACAGCGGTGCGACATGGCGAGATTCGATTCCG 1050
 :::
 334 IeThrAlaHisProIleLysArgSerGlnMetGlyAlaIleAlaLeuPro 350
 1051 AAGGGGAATCCGCGTCACGCGCATTTTGGCGGATGGCGATACGCCAA 1100
 :::
 351 LysGlyLysSerAlaValSerAspAsnPheAlaAspAlaIleTyrAlaLys 367
 1101 ATACCGTCCCTTACCATTCGCGAAATATCCGTTCAACTTGGAGCAGC 1150
 :::
 367 sTyrProSerProTyrHisSerArgAsnIleArgSerAsnLeuGlnGlnA 384
 1151 GTTACGCGCAAAACAAACATCACTGCTCAACGCTGCGCGGCGTCAACGGA 1200
 :::
 384 rgtYrGlyLysGluAsnIleThrSerThrValProProSerAsnGly 400
 1201 AAGAAATGTAACTGGCAACAAACGCAACCGGCAAGCAAAATGCGCTT 1250
 :::
 401 LysAsnValLysLeuAlaAspGlnArgHisProLysThrGlyValProPh 417
 1251 TGAACGTAAGGTTTCCGATTTGAAAACGTAATAATACGATACGA 1300
 :::
 417 eaSpGlyLysGlyPheProAsnPheGlnLysHisValLysTyrAspPheL 434
 1301 GAATTAATACCGCTGTAACCAACGATGTAAT.....CCTATAGAT 1338
 :::
 434 yLysLeuAsp.....IleGlnGlnLeuSerIleGlyLysIleProLysAla 448
 1339 GAACCGCTCTTAAATCCTAAAGGTTCTGTCGATCGGCTCATTTCTTGTC 1388
 :::
 449 LysProValPheAspAlaLys.....ProArgTyrPheI 459
 1389 TATTAATGCCAGAAATTCATACGCAAAATATACCAAGCAAGATGAGATCA 1438
 :::
 459 uValAspArgLysLeuAsnLysLeuThrThrArgGlnGlnValGlyLys 475


```

307 AlaGluThrValGluAlaLeuValAsn.....ValLeuProPh 319
957 CAAGGAGATTGAGCTGTTCGGGAAATACGGCTTGGCGGCATCAGCAG 1006
319 e..... 319
1007 CACATCCGTGTCAAGCGGTCCGAGATGGCGGAGATCGATTCGGAAAGG 1056
320 .....AlaLysValLysAsnLeuThrLysAlaAlaLysPro...Gly 332
1057 AAATCCGCGCGTACGAGCAATTTGGCATCGGCGATACGGCAATACCC 1106
333 LysAlaAlaValSerGlyAspSerAlaAla..... 343
1107 GTCCCTTACCATTCGCGAAATACCT..... 1134
344 .....TyrAsnThrArgThrThrArgLysValThrThrGluGlu 358
1135 .....TCAACTTG...GACGAG 1149
358 LLeuAsnArgLLeaArgLAsnGlnLysAsnSerAsnLLeuGluLys 374
1150 CGTTACGCGCAAGAAACATACCCCTCCACGCTGCCGCGCTCAAC.. 1197
375 AsnTyrGlyArgAspAsnProAsnHisLLeaValLeuSerGlyAsnSe 391
1197 ..... 1197
391 rLeGlnHisLLeuTyrGlyAspGluAlaGlyGlyGlyHisLeuPhe 408
1198 ..GGAAAGATGTGAACGTGGCAACAAACGCGC.. 1230
408 roGlyLysProGlyLysThrPheProGlnHisTrpSerAlaSerLys 424
1231 .....CCGAAGACCAAA..... 1242
425 lIleThrHisGluLeuSerAspLLeuValThrSerProLysThrGlnTrpTy 441
1243 .....GTCCGTTTGAAGGTTCGCGAATT 1273
441 rAlaGlnThrGlyThrGlyLysTyrLLeuAlaLysGlyArgPro.... 456
1274 TTGAAGAAAGCAATTAATACGATACGGA..... 1302
457 ..AlaArgTPrValSerTyrGluThrArgAspGlyLLeaArgThr 472
1303 .....ATTAAATACGCGCTGTACCAAGCAAGCA 1328
473 ValTyrGluProAlaThrGlyLysValThrAlaPhePro.....As 487
1329 TCCATATGATGACCCGCTTTAATCT 1356
487 pArgThrSerAsnProLysTyrAsnPro 496

seq_name: sp_bacteriap:09KOF4

seq_documentation_block:
ID 09KOF4 PRELIMINARY; PRT; 422 AA.
AC 09KOF4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NMRP-RELATED PROTEIN.
GN NMR0653.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCB1_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RA MEDLINE=20175753; PubMed=10710307;
RA Tetrelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,

```

```

RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gali J., Scariato V., Masianni V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002420; AAF62314.1; -.
DR TIGR; NMR0653; -.
KW Complete proteome.
SQ SEQUENCE 422 AA; 45845 MW; B5428F16A296585F CRC64;

```

```

alignment_scores:
  Quality: 662.50      Length: 461
  Ratio: 2.427        Gaps: 11
  Percent Similarity: 59.219  Percent Identity: 33.623

```

alignment_block:

US-09-303-518d-465 x 09KOF4

Align seg 1/1 to: 09KOF4 from: 1 to: 422

```

13 CGCAAAATATCCCTTATTCG.....TCATATCTGCAGTGTGCT 53
|||||..... 5
ArgArgLeuThrAsnLeuAlaLysAlaValAlaAlaAla 21
54 GCCGATGCATGCACAGCCTCAGATTTGGCAAGCTTTCTTATTCGGC 103
21 uLleGlnProAlaLeuAlaAlaAspLeuAlaGlnAspProPheIleTha 38
104 AGCTTTCGACGCTCAGCATTTTCGAACCGGAGGAATATCAGCTATTC 153
38 spAsnAlaGlnArgGlnHisTyrGluProGlyGlyLysThrHisLeuPhe 54
154 GCGAGC..AGGGGGAAGTGGCGGCGGCGGCGATATGCGATGGG 200
55 GlyAspProArgLysValSerAspArgThrGlyLysLLeaValAl 71
201 AAACATATCAAGCATCAGTTGGGCAACCTGTTCATCCAGCGGCGCA 250
71 eGlnAspTyrThrHisGlnMetGlyAsnLeuLeuLleGlnAlaAsnI 88
251 TTAAGGAATATCGCTACATTTGCCGTTTCCGATCCGCGGCA 300
88 lAsnGlyThrLleGlyTyrHisThrArgPheSerGlyHisGlyHisGlu 104
301 GTCCATTCGCCCTTCGACACCATGCTCACAATTCGATTCGATGAAGC 350
105 GlnHisAlaLProPheAspAsnHisAlaAlaAspSerAlaSerGluGlu 121
351 CGGTAGTCCGTTGACGAGATTACGCTTACCGCATTCATTCGAGGAGAT 400
121 sGlyAsnValAspGluGlyPheThrValTyrArgLeuAsnTrpGluGln 138
401 AGCAACACCATCCCGCGGCGGCTATGACGGGCGGCGGCGGCTAT 450
138 lSGlnHisLProAlaAspAlaTyrAspGlyProLysGlyGlyAsnTyr 154
451 CCCGCTCCCAAGGCGCGAGGATATATACAGCTACGATCAAAAGGCGT 500
155 ProLysProThrGlyAlaArgAspGluTyrThrThrHisValAsnGlyTh 171
501 TGCCCAAAATATCCGCTCAACCTGACGCGACACCGACGACAC 550
171 lAlaArgSerLLeuLysLeuAsnProThrAspThrArgSerLLeaArgLna 188
551 GCGTGTGCGAGTTTCCACAATACCGGTAGTATGCTACCAAGAGTA 600
188 rGlnLeuSerAspAsnTyrSerAsnLeuGlySerAsnAspPheSerAspArgAla 204

```

```

601 GCGCAGCGATTCAAAACCGCCGACCGGATACAGCCCGGAGCTGACAGATC 650
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
205 AspGluAlaAsnArgLysMetPheGlnHisAsnAlaLysLeuAspArgTr 221
      ||||| : : : : : : : : : : : : : : : : : : : : : :
651 GGGCAATAGCCGCGAGGATTTCAACGCGACCTGCGATATGCTCAAAAACA 700
      ||||| : : : : : : : : : : : : : : : : : : : : : :
221 pGlyAsnSerMetGluPheIleAsnGlyValAla..... 232
      ||||| : : : : : : : : : : : : : : : : : : : : : :
701 TCATCGCGCGCGGAGAGAAATTTGCGCGCAGCGCATGCCGTCGACAGGT 750
      : : : : : : : : : : : : : : : : : : : : : :
233 ..AlaGlyAlaLeuAsnProPheIleSerAlaGlyGluAlaValAsp... 247
      : : : : : : : : : : : : : : : : : : : : : :
751 ATAAGCGAAGGCTCAACATGCTGTATGACGCGCTTGGGCTGCTTTC 800
      : : : : : : : : : : : : : : : : : : : : : :
247 ..... 247
801 CACGCAAAACAAGATGGCGCGCATCAACGATTGGCAGATATGGCGCAAC 850
      : : : : : : : : : : : : : : : : : : : : : :
247 ..... 247
851 TCAAAGACTATCCGCGACAGCCATCCGCGATTTGGGCGATCCAAAACCCC 900
      : : : : : : : : : : : : : : : : : : : : : :
248 .....GlnTrpMetGlnGluAsnPro 254
      : : : : : : : : : : : : : : : : : : : : : :
901 AATGCCGCAAGCATAGAGAACCGCGTACAGAAATATCTTTACGCGCATCAT 950
      ||||| : : : : : : : : : : : : : : : : : : : : : :
255 AsnAlaAlaGluThrIvalGluAlaLeuValAsn.....ValLe 267
      : : : : : : : : : : : : : : : : : : : : : :
951 CCCCCTCAAGGGATGGAGCTGTTCGGGAAATACGCGCTTGGGCGGCA 1000
      : : : : : : : : : : : : : : : : : : : : : :
267 uProPhe..... 269
1001 TCACGCGCATCTCTGTAAGCGGTCGAGATGGGCGAGATGTCATTCGCG 1050
      : : : : : : : : : : : : : : : : : : : : : :
270 .....AlaLysValLysAsnLeuThrLysAlaAlaLysPro 281
      : : : : : : : : : : : : : : : : : : : : : :
1051 AAAGGAAATCCGCGCGTACGAGCAATTTTCCGATGGCGCATACGCCAA 1100
      ||||| : : : : : : : : : : : : : : : : : : : : : :
282 ...GlyLysAlaAlaValSerGlyAspPheSerAspSer..... 293
      : : : : : : : : : : : : : : : : : : : : : :
1101 ATACCCGTCCTTACCATTCGCGGAAATATCCGTTCAACTTGGACGACG 1150
      : : : : : : : : : : : : : : : : : : : : : :
294 .....TyrLysHisAsnThrAlaSerArgLeuSerGlnS 305
      : : : : : : : : : : : : : : : : : : : : : :
1151 GTTACGCGCAAGAAACAATCACTCTCCATACCGCTGCGCGCTCAAAACGA 1200
      ||| : : : : : : : : : : : : : : : : : : : : : :
305 eValAlaSpGlyGluMetPheGlnThrArgAsnVal.....AspPhe 318
      : : : : : : : : : : : : : : : : : : : : : :
1201 AAGAAATGTGAACCTGGCAAAACAGCCGCAAGACCAAAAGTGGCGGT 1250
      : : : : : : : : : : : : : : : : : : : : : :
319 LysAlaLysSerIleGlyThrLysIleHisAspGlyAlaGln..... 332
      : : : : : : : : : : : : : : : : : : : : : :
1251 TCACGCGTAAAGGTTCCGAATTTGAAAGAAGAGCTAAATATACGATAGA 1300
      ||||| : : : : : : : : : : : : : : : : : : : : : :
333 ....GlyLysHisIleSerGlyHisArgAsnLysTrIleGluGlyLysSer 348
      : : : : : : : : : : : : : : : : : : : : : :
1301 GAATTAATACCGCTGTACCAACAAGTAATCTATAGATGAACCGCTCTTT 1350
      : : : : : : : : : : : : : : : : : : : : : :
348 hLeuAsnGlnAsnIle..... 353
      : : : : : : : : : : : : : : : : : : : : : :
1351 AATCTTAAAGTTTCTGCGATCGGCTATCT 1383
      ||||| : : : : : : : : : : : : : : : : : : : : : :
354 AsnProGlnGluLeuLeuAsnGlyLysIleHisSer 364
      : : : : : : : : : : : : : : : : : : : : : :
seq_name: sp_bacteria:Q9X6P3
seq_documentation_block:
ID Q9X6P3 PRELIMINARY; PRT; 509 AA.
AC Q9X6P3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ADHESIN MAFB.

```

```

GN MAFB.
OC Neisseria gonorrhoeae.
OC Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OK NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11;
RA Eickernjaeger S., Meyer T.F., Fischer E., Maier J., Manning P.A.,
  Rudel T., Scheuerpflug I., Schulz E., Schwan E.T.;
  Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF142582; AAD31039.1; -; 88BA20CEFD8269D4 CRC64;
SQ SEQUENCE 509 AA; 55681 MW; 88BA20CEFD8269D4 CRC64;

alignment_scores:
  quality: 454.00      Length: 411
  ratio: 1.884        Gaps: 18
  Percent Similarity: 58.637      Percent Identity: 32.847

alignment_block:
US-09-303-518D-465 x Q9X6P3 ..
Align seg 1/1 to: Q9X6P3 from: 1 to: 509

31 CTGTCCATATCTGGCAGTGTGCTGCGGATGATGCACAGCGCTCAGATT 80
   : : : : : : : : : : : : : : : : : : : : : :
16 IleSerLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
   ||||| : : : : : : : : : : : : : : : : : : : : : :
81 GGCAAACGATCTTTATCCGCGAGGTTCTCGACCGCTCAGCATTTGCAAC 130
   | : : : : : : : : : : : : : : : : : : : : : :
29 u.....AspAlaArgLeuArgAspAspMetGlnAlaLysHisIstYrGluP 44
   ||||| : : : : : : : : : : : : : : : : : : : : : :
131 CCGACGCGCAAAATACACCATATTCGCG..AGCAGGGGCGAATTCGCGAG 177
   || : : : : : : : : : : : : : : : : : : : : : :
44 roGlyGlyLysTyrHisLeuPheGlyAsnAlaArgGlySerValLysAsn 60
   : : : : : : : : : : : : : : : : : : : : : :
178 CGCAGCGGTTCATTCGCGATTGGGAATACAAAGCGATCAGTGGGCA 227
   || : : : : : : : : : : : : : : : : : : : : : :
61 ArgVal.....CysAlaValGlnTrpPheAspAlaThrAlaValGlyTr 75
   : : : : : : : : : : : : : : : : : : : : : :
228 CTGTTC.....ATCCAGCAGCGCGCATTTAAAGGAAATATTCGCT 268
   : : : : : : : : : : : : : : : : : : : : : :
75 oIleLeuProIleThrHisGluArgThrGlyPheGluGlyIleIleGlyT 92
   || : : : : : : : : : : : : : : : : : : : : : :
269 ACATTTGCGCTTTTCCGATCCGCGCAGCAAGTTCATTCGCTTCGAC 318
   || : : : : : : : : : : : : : : : : : : : : : :
92 yrcGluThrHisPheSerGlyHisGlyHisGlyValHisSerProPheAsp 108
   : : : : : : : : : : : : : : : : : : : : : :
319 AACCATGCCCTCACATTCGCGATTCGATGAAGCGGTAATCCGTTGACGG 368
   ||||| : : : : : : : : : : : : : : : : : : : : : :
109 AsnHisAspSerLysSerThrSerAspPheSerGlyGlyValAspGlyLe 125
   : : : : : : : : : : : : : : : : : : : : : :
369 ATTACAGCTTTACCGCATCCATTTGGAGCGGATACGACACCATCCGCGG 418
   ||||| : : : : : : : : : : : : : : : : : : : : : :
125 yPheThrValTyrGlnLeuHisArgThrGlySerGluIleHisProAla 142
   : : : : : : : : : : : : : : : : : : : : : :
419 ACGGCTATGACGGGCGACAGGGGCGGCTATCCGCTCCGAAAGGCGGG 468
   ||||| : : : : : : : : : : : : : : : : : : : : : :
142 spGlyTyrAspGlyProGlnGlyGlyLysTyrProGluProGlnGlyAla 158
   : : : : : : : : : : : : : : : : : : : : : :
469 AGGAGTTATATACAGCTACGATTAAGAGCGGTGCCAAATATTCGCGCT 518
   ||||| : : : : : : : : : : : : : : : : : : : : : :
159 ArgAspIleLysSerTyrHisIleLysGlyThrSerThrLysThrLysI 175
   : : : : : : : : : : : : : : : : : : : : : :
519 CAAC.....CTGACCGCACACC 535
   ||||| : : : : : : : : : : : : : : : : : : : : : :
175 eAsnThrValProGlnAlaIaProPheSerAspArgTrPleuLysGluAsn 192
   : : : : : : : : : : : : : : : : : : : : : :
536 GCAAGCAGGAGCAACAGGCTTGTGACCGCTTCCACAAATACCGGTACTAG 585
   : : : : : : : : : : : : : : : : : : : : : :
192 IaglyAlaAlaSerGlyPheLeuSerArgAlaAspGluAlaGlyLysLeu 208
   : : : : : : : : : : : : : : : : : : : : : :
586 CTGACGCAAGAGTAGGCGACGATTCAAACGCGCGCATACCGGATACAGCC 635

```

RA	Whitehead S., Spratt B.G., Barrell B.G.:		
RT	"Complete DNA sequence of a serogroup A strain of <i>Neisseria meningitidis</i> Z2491."		
RL	Nature 404:502-506(2000).		
DR	EMBL; AL162752; CAB83629.1; ..		
KW	Complete proteome.		
SV	SEQUENCE 482 AA; 52339 MW; 805C99EB86ED3DA CRC64;		

Alignment_scores:			
Quality:	419.50	Length:	447
Ratio:	1.665	Gaps:	18
Percent Similarity:	56.376	Percent Identity:	31.320

alignment_block:			
US-09-303-51BD-465 x Q9JWK8	..		

Align seg 1/1 to: Q9JWK8	from: 1 to: 482		
--------------------------	-----------------	--	--

31	CTGTCCATCTAGCGACATGTGCGTCGGCGGATGCATGCACACCCCTCAGATT	80	
16	lIeSerLeuGlnIleProIle.....SerHisAlaSnGly	29	
81	GGCAAAACAGTTCTTTATCCGGCAGGCTTCGCACCTCAGATTTGCAAC	130	
29	l.....AspAlaArgLeuArgAspAspMetGlnAlaIleValysIleGlu	44	
131	CGAGCGGAAATACCACTATTCCGC..AGCAGGGGGAACTTGGCCGAG	177	
44	toGlyGlyLysTrpHisLeuPheGlyAsnAlaArgLysSerValysAsn	60	
178	CGCAGCGGTCAATATGGATTGGTGGAAACATACAAAGCCATCAGTTGGGCA	227	
61	ArgVal.....TyrAlaValGlnThrPheAspAlaThrAlaValGly	75	
228	CCTGTTCT.....ATCCAGCAGGGCGGCATTAAAGAAATATCGCT	268	
75	oIleLeuProIleThrIleHisAlaArgThrGlyPheGlnGlyIleIleGly	92	
269	ACATGTCGCGTTCCTCCGATCGAGGGCAGAGTCATCCGCCCTTCGAC	318	
92	YrGluThrHisPheSerGlyHisGlyHisGlyValHisSerProPheAsp	108	
319	AACCATGCCATCAATTCGGATTCTGATGAAAGCGGATAGCCGTTGACGG	368	
109	AsnHisAspSerLysSerThrSerAspPheSerGlyGlyValAspGly	125	
369	ATTACGCTTTACGCATCCATCTGGGAGCGAGTAAACACATCCCGCGG	418	
125	YpHeThrValTyrGlnIleuHisArgThrGlySerGlnIleHisProGlu	142	
419	ACGGTATGACGGGCGCACAGGGCGGCTATCCCGCTCCCAAGGGCGG	468	
142	spGlyTyrAspGlyProIleGlnSerAspTyrProProIleGlyAla	158	
469	AGGATATATACAGTACAGATAAAGAGCGTTGCCAAATATCCGCT	518	
159	ArgAspIleTyrSerTyrTyrValysGlyThrSerThrLysThrLys	175	
519	CAACCTGACCGCAACCGCACCGGACGCAACGCTTTCGACCGTTTCC	568	
175	TasIleValProArgAlaProPheSerAspArgTyrLeuLysGluAsn	192	
569	ACAATACCGGTAGTATCTGTGACGAGGAGTAGGAGCAGGATTCAAAGC	618	
192	IaGlyAlaIleAspSerGlyPhePheSerArgAlaAspGlnIleGlyLys	208	
619	GCCACCGGATACACGCCGAGCTGGACAGATGGGCAATGCCCGGCAAG	668	
209	lIeTrpIleuSerAspProAsnLysAsnTyrTrpAlaSerArgMetAsp	225	
669	TTTCAACGGCAGCTGCATATTCGCMAAAC.....ATCATCGGCGGG	712	

```

225 pIleatgglYlIleValGInclYAlaValAsnProPheleuMetGlyPheG 242
713 CAGAGAAATTTGTGGCGCA.....GGCGATGCGTGCAGGGTATAGC 756
242 InqIValaIglYlIleGlyAlaIleThrAspSerAlaValSerProValThr 258
757 GAAGGCTCAACATTCGTCTTATGCAAGCG.....TTGGCTGTGCTG 797
259 AspThrAlaAlaIleInqIleuInqIleuInqIleuInqIleuInqIleu 275
798 TTCACCGCAAAACAGATGGCGGCATCAAGATTTGGCAGATATGCG. 846
275 UserProGluAlaGlnLeuAlaIleThrAlaLeuAlaAspSerAlaP 292
847 ..CAACTCAAGACTATGCGGACAGCCATCGCGATTTGGCAGTCCAA 894
292 heaIaValIysAspGlyIleAsnSerAla...ArgGlnTrpAlaAspAla 307
895 AACCCCAAT.....GGCGACAGGCGATGAAAGCCGTCAGACAAATAT 935
308 HisProAsnIleThrAlaThrAlaIleGlnThrAlaLeuAlaGluAl 324
936 CTTTACGCGACATCATCC..... 954
324 aaIaThrThrValTrpGlyIleValIleValIleValAsnProThrLys 341
955 .....GTCAAGGATTTGA.....GCTGTTCGGGGGAAA 984
341 rPAspTrpValIysAsnThrGlyTrpLysThrProAlaValArg..... 355
985 TACGGCTTGGGCGCATCAGCGCATCTGTCAAGCGGTCGCGAGATGG 1034
356 .....ThMetHisThrLeuAspGlyIleMetAlaGlu 366
1035 CGAGATCGCATTCGCCGAAGGAAATCCGCGTCAGGACATTTTCCG 1084
366 yGIYAsnArgProPro.....LysSerIleThrSerAsnSerLysAla 381
1085 ATGGG..... 1089
381 sPAlaSerThrGlnProSerIleuAlaGlnIleuIleGlyGluGlnIle 397
1090 .....GCATACGCCCAATATC.....CCGTCCCTTACCA 1118
398 SerSerGlyHisAlaTrpAsnLysHisValIleArgGlnGlnGluPheTh 414
1119 TTCGCGAAATTCGCTCAAACTTGACAGCAGCTTACGCGCAAGAAACA 1168
414 rAspLeuAsnIleAsnSerProAlaAspPheAlaArgHisIleGluAsn 431
1169 TCACCTCTCAACCGTCGCCGCTCAACGGAAGGAATGTG 1209
431 IeValSerHis.....ProThrAsnMetLysGluLeu 441

seq_name: sp_bacteriap:09JXD2

seq_documentation_block:
ID 09JXD2 PRELIMINARY; PRT; 483 AA.
AC 09JXD2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MAFB PROTEIN.
GN MMB2105.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,

```

```

RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Cleek A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Uettersack T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RT Science 287:1809-1815(2000).
DR EMBL: A802559; AAF62340.1;
DR TIGR: MMB2105;
KW Complete proteome.
SQ SQUENCE 483 AA; 52789 MW; D6B5AFRA8A8BA02 CRC64;

```

alignment_scores:

```

Quality: 404.50 Length: 553
Ratio: 1.385 Gaps: 25
Percent Similarity: 52.803 Percent Identity: 26.944

```

alignment_block:

US-09-303-518D-465 x 09JXD2 ..

Align seg 1/1 to: 09JXD2 from: 1 to: 483

```

31 CTTGTCATATGTCAGATGTCCTGCCATGCATGCACACGCTCAGATT 80
:::|||||:::  :::  :::|||||:::  ||
16 ILeSerLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
81 GCGCAACGATTTCTTTATCCGCGAGTTTCGACGCGTACGATTTGCAAC 130
29 1.....AspAlaArgLeuArgAspAspMetGlnAlaLysHisTyrGluP 44
131 CCGACGCGAAATATCCACSTATTTCGCG...AGCAGGGGGGAACTTGCAG 177
11  |||||||  :::|||||  :::|||||  :::  :::
44 rGclYglYstYrGlnHisLeuPheGlyAsnAlaArgLysSerValLysLys 60
178 CGCAGCGGTCATATCGATTTGGGAAACATCAACACCATCAG..... 219
61 Arg.....ValTyrAlaValGlnThrPheAspAlaThrAl 72
220 .....TTGGCAACCTGTTCATCCAGACAGCGGCATTAAGGA 259
72 aValSerProValLeuProIleThrHisGluArgThrGlyPheGluGly 89
260 ATATCGGCTACATTTGCGCTTTCCGATACAGGCGCAGACATTCATCC 309
89  |||||||  :::|||||  |||||||  |||||||  |||||||
89 aIleGlyTyrGlnThrHisPheSerGlyHisGlyHisGlyValHisHis 105
310 CCGTTGACACACCATGCTCAGATTCGATTCGATGACAGCGGATGTCG 359
106 ProPheAspHisHisAspSerLysSerThrSerAspPheSerGlyGly 122
360 CGTTGACGATTCAGCTTTACCGCATTCATTCGATGACAGCGATACACC 409
122 IAspGlyGlyPheThrValTyrGlnLeuHisAspGlyThrGlySerGlu 139
410 ATCCGCGCGGCGCTATGACGCGGCGCAGGCGGCGGCTATCCGCTCC 459
139 IAspGlyAspGlyTyrAspGlyProGlnLysSerAspTyrProPro 155
460 AAAGGCGCGAGGATATATACGCTACGATCAAAAGCGGTTGCCCAAAA 509
156 GLyGlyAlaArgAspIleTyrSerTyrTyrValLysGlyThrSerThr 172
510 TATCCGCTCACCCTGACGCGACACCGCAGACCGGACCAACGGCTTG 559
172 sThrLysThrAsnIleValProGlnAlaProPheSerAspArgTyrPleu 189
560 ACGGTTCCACATATACCGTAGTATGTCGACGCAAGAGATGAGCGAG 609
189 ysgIuAsnAlaGlyAlaHisSerGlyPhePheSerArgAlaAspGluAla 205

```



```

610 TTCAAAGCGCCACCGATACAGCCCGAGCTGACAGATCGGGCAATGC 659
    |||      |||      |||      |||      |||      |||
206 GLYLeuLeuIleThrIleuSerProAsnIleThrIleAsnIleThr 222
    |||      |||      |||      |||      |||      |||
660 CGCCGAAGCTTTCACAGCGACTGACATATCGTCAAAAC.....ATCA 703
    |||      |||      |||      |||      |||      |||
222 GMetAspAspValArgIleValAlaGlnIleValAlaValAsnProPheLeu 239
    |||      |||      |||      |||      |||      |||
704 TCAGCGCGGAGAGAAATTTGCGGCA.....GGCATCCGTGCAG 747
    |||      |||      |||      |||      |||      |||
239 euclIleGlnIleValIleGlyAlaIleThrAspSerAlaValSer 255
    |||      |||      |||      |||      |||      |||
748 GGTATACGGAAGGCTCAACATGCTGTATGACAGGC.....TT 788
    |||      |||      |||      |||      |||      |||
256 ProValThrAspThrAlaIleGlnIleGlnIleGlnIleGlnIleAsnAspLe 272
    |||      |||      |||      |||      |||      |||
789 GGGTCTGCTTTCACCGAAACACATGCGCGCATCAACGATTTGGCAG 838
    |||      |||      |||      |||      |||      |||
272 uclIleuSerProGluAlaGlnLeuAlaAlaIleSerLeuGlnA 289
    |||      |||      |||      |||      |||      |||
839 ATATGGCG...CACTCAAGACATATGCGCGAGCCATCGCGATGG 885
    |||      |||      |||      |||      |||      |||
289 spSerAlaPheAlaValIleAspGlyIleAsnSerAla...LysGlnIleP 304
    |||      |||      |||      |||      |||      |||
886 GCAGTCCAAACCCCAATGCCGACACAGGATAGAACGCTCAGCATAT 935
    |||      |||      |||      |||      |||      |||
305 AlaAspAlaIleProAsn.....IleThrAlaIleThrAlaGln.. 316
    |||      |||      |||      |||      |||      |||
936 CTTTACGGAGTCATCCCGTCAAGGATGAGCTGT...CGGGAA 982
    |||      |||      |||      |||      |||      |||
317 ...ThrIleuSerAlaIleAlaIleAlaIleGlyThrValIlePArgGlyL 332
    |||      |||      |||      |||      |||      |||
983 AATACGCGCTGGGCGGATACAGGACATCTGTCAACGCTCGCAGATG 1032
    |||      |||      |||      |||      |||      |||
332 ys.....LysValGlnLeuAsnProThrLys..... 340
    |||      |||      |||      |||      |||      |||
1033 GCGGAGATCGATTCGCGAAAGGAAATCCCGCGACGACGCAATTTGC 1082
    |||      |||      |||      |||      |||      |||
341 .....TTPAspTrpValLys 345
    |||      |||      |||      |||      |||      |||
1083 CGATGCGGATACGCAATACCGCTCCCTTACCATTCGCGAAATATCC 1132
    |||      |||      |||      |||      |||      |||
345 sasnThrIleGlyTrpLysLys...ProAlaIleAlaArgHisMetClnThrIleuA 361
    |||      |||      |||      |||      |||      |||
1133 GTTCAAACTTGAGCAGCGTTACGCGCAAGAAACATACCTCTCAACC 1182
    |||      |||      |||      |||      |||      |||
361 spGly.....GlnMetAlaGlyIleLysnLysProIleLysSer..... 373
    |||      |||      |||      |||      |||      |||
1183 GTGCCGCGCTCAACGCAAGATGTGAATCTGCAACAAACGCGCACCC 1232
    |||      |||      |||      |||      |||      |||
374 .....LeuProAsnSerAlaIleAlaGln 380
    |||      |||      |||      |||      |||      |||
1233 GAAGACCAAGTGCCTTTGACGGTAAAGGTTTCCGATTTTGAATAA 1281
    |||      |||      |||      |||      |||      |||
380 uLysArgLys.....GlnAsnPheGlnIleLysP 389
    |||      |||      |||      |||      |||      |||
1282 .....GACGTAAATACATACGACATTAATATACCGCT 1314
    |||      |||      |||      |||      |||      |||
389 heanSerAsnTrpSerSerAlaSerPheAspSerValHisLysThrLeu 405
    |||      |||      |||      |||      |||      |||
1315 GTACACCAAGTGAATCTTATAGATGAACCGCTTTAATC..... 1354
    |||      |||      |||      |||      |||      |||
406 ThrProAsnAlaProGlyIleLeuSerProAspLysValLysThrArgT 422
    |||      |||      |||      |||      |||      |||
1355 .....CTAAGGTTGTGCGATCGGCTCATTTCTT 1384
    |||      |||      |||      |||      |||      |||
422 yThrSerLeuAspGlyLysIleThrIleIleLysAsp..... 434
    |||      |||      |||      |||      |||      |||
1385 GGTCTAATACGCGCAATATACGCAAAATTAAC..... 1420
    |||      |||      |||      |||      |||      |||
435 .....AsnGluAsnAsnIleTrpPheArgIleHisAs 444
    |||      |||      |||      |||      |||      |||
1421 .....CAAGCAAGGTAGATCAGATATATCCACCTAAATAATTA 1460
    |||      |||      |||      |||      |||      |||

```

```

444 pAsnSerArgLysGlnIleuAspSerAsnGlyAsnAlaValLysThrG 461
    |||      |||      |||      |||      |||      |||
1461 CTCCTCTTACGACCGCGCTACCAAAAGACCTTAATATGATTTGGATA 1510
    |||      |||      |||      |||      |||      |||
461 LysnLeuGlnIleLysGlnAlaLysAspTrpLeu..... 472
    |||      |||      |||      |||      |||      |||
1511 AATTTGTAATGAATGACACTAAAGTCCATCAACAGCACTAAAGTCAAGA 1560
    |||      |||      |||      |||      |||      |||
473 .....GlnGlnIleThrHisIleArgAs 480
    |||      |||      |||      |||      |||      |||
1561 TTTCGAT 1567
    |||      |||      |||      |||      |||      |||
480 nLeuAsp 482
    |||      |||      |||      |||      |||      |||

seq_name: sp_human:09U039
seq_documentation_block:
ID 09U039 PRELIMINARY; PRT; 956 AA.
AC 09U039;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010088; BA083714.1; -.
DR IntepPro; IPR002965; P_Rich_extensions.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
FT NON_TER 956
FT SEQUENCE 956 AA; 106462 MW; C9FEAD07B49F5011 CRC64;

alignment_scores:
Quality: 173.50 Length: 537
Ratio: 0.700 Gaps: 27
Percent Similarity: 46.182 Percent Identity: 24.209

alignment_block:
US-09-303-518d-465 x 09U039 ..
Align seg 1/1 to: 09U039 from: 1 to: 956

35 CCATCTGGCAGTGTGCTGCGATGACGACAGCCCTCAGATTTGGCA 84
    |||      |||      |||      |||      |||      |||
476 ProGlnIleP.....ArgArgSerArgSerAlaGlnArgTrp.. 487
    |||      |||      |||      |||      |||      |||
85 AACGATCTTTATCCGCGAGGTTCTGCACCGTTCAGCATTTGCAACCCGA 134
    |||      |||      |||      |||      |||      |||
488 .....GlyArg.SerArgSerProGlnArgArgGlyArg 498
    |||      |||      |||      |||      |||      |||
135 CGGGAATATCCACCTATTTGCGCAGCAGGGGGAACCTTGCAGGCGCAGC 184
    |||      |||      |||      |||      |||      |||
499 SerArgSerPro.....GlnArgProGlyTrpSerArgSerArgAs 512
    |||      |||      |||      |||      |||      |||
185 GTCAATTCGATTCGGAACATACAAAGCCATCACTTGGCACTGTTC 234
    |||      |||      |||      |||      |||      |||
512 nThrGlnArgArgGlyArgSerArgSerAla..... 522
    |||      |||      |||      |||      |||      |||
235 ATCCAGCAGCGCGCATTAAGGAATATCGGCTCAT.....TGTCCG 278
    |||      |||      |||      |||      |||      |||
523 .....ArgArgGlyArgSerHisSerArgSerPro 532
    |||      |||      |||      |||      |||      |||
279 CTTTCCGATTCACGGGCAAGATGTCATTTCCCTTCGACAAACCATGCGCT 328
    |||      |||      |||      |||      |||      |||

```

```

533 AlAThArgIyArGSerArGThrProAlaArGArg..... 546
329 CAATTCOCGATTCGTGAAGCGGAGTCCGCTGACGATTCAGCTT 378
547 .....GlyArGSerArGSerArGThr 554
379 TACCGATTCATTGGAGGATACGAACACATCCGCGGAGCTATGCA 428
554 roAlaArGArGSerArGThrProThrArGArGArGSerArG 570
429 CGGGCCAGAGGGGGGCTATCCCGCTCCCAAGGGCGAGGATATAT 478
571 SerArGThrProAlaArGArGArGSerArGThrPro...Al 586
479 ACAGCTACGACATAAAGCGCTTGCCTCAATTCGCTCAACCTGACC 528
586 aArGArGArGSerArGThrArGSerProValArGArGSerArG 603
529 GACACCGCAGACCGGACACAGCGCTTGTGACCG...TTTCACATAC 575
603 rGSerProAlaArGArGSerGlyArGSerArGThrProAlaArG 619
576 CGGTAGTATGCTGACGACAGAGAT.....AGCGGACGATTCAC 616
620 ArGArGArGSerArGSerArGThrProAlaArGArGArGSerArG 636
617 GCGCCACCGCATACAGCCCGAGCTGACAGATGGGCAATGCCGC...C 663
636 rArGThrPro.....AlaArGArGSerGlyArGSerArG 649
664 GAAGCTTTCACAGGACGCTGACATATGCTCAAAACATCATCGCGCGG 713
649 rGThrProAlaArGArGArGSerArGThrProArGArG 665
714 AGGAAATTCGCG.....CGACGCGATCCGCTGC 745
666 ArGSerArGSerArGSerLeuValArGArGArGSerArGThr 682
746 AGGATATAGGAAAGCTCAACATTCGCTTATGCAAGCGCTTGGCTG 795
682 rProGlnArGArg.....GlyArGSerArG 692
796 CTTTCACCGGAAA...CAAGATGGCGGATCAACGATTTGGAGATAT 842
692 erSerGlnArGArGArGSerArGThrSerGlnArGArGSerArG 708
843 GCGGCACTCAAGACTATGCCGCGAGC.....AGCCATCCGCGA... 881
709 AnSerSerProGlnMetLysSerArGLeuSerArGThrArG 725
882 ....TTGGGAGTCCAAACCCCAATGCCGACAGGATAGAGCCGCTC 927
725 gSerLeuSerSerProArGSerArGLeuLysSerArg...LeuSerLeuA 741
928 AGCAATATCTTACGGCAGTATCC..... 953
741 rGArGSerLeuSerGlySerSerProCysProLysGlnLysSerGlnThr 757
954 .....CGTCAAGGATTCGAGCTGTTGGGGAATAACGGCTTGG 994
758 ProProAlaArGArGSerArGLeuLysSerGlnProLysAlaLys.... 772
995 GCGGATCAGCGGACATCCTGTCAAGCGGTGCGAGATGGGAGATCCGA 1044
773 .....SerArGThr.ProProArGArGSerArGSerSerSerPro 786
1045 TTGCGGAAGGGAATCCCGCTCAGCGCAATTTTGGCGATGGGCAATA 1094
787 ProProLysGlnLysSerArGThrProSerArGArGLeuSerSerSe 803
1095 CGGCAATATACCGCTCCCTTACCATTCGCAATATCCGCTTCAAACTGG 1144
803 rSerProHisPro.....LysValLysSerGlyThrP 814

```

```

1145 AGCAGCTTACGGCAAGAAACATACCTTCACCGTCCGCGCTCA 1194
814 roProArGInGly.....SerLeuThrSerProGlnAla 825
1195 AACGGAAGAATGTGAACCTGCAACAAACAGCGCAGGACCAAGT 1244
826 AsnGlnGlnSerValThr.....ProGlnArGArGSe 836
1245 GCCCTTACCGGTAAAGGTTTCCGAATTTGAAAAGAGCTA..... 1287
836 rCysPheGlnSerSerProAspProGlnLeuLysSerArGThrPro 853
1288 .....AAATACGATACGAGATTAATACCGCTGTACCA 1320
853 rGHisSerCysSerLysSerSerProProArGValLysSerSerThrPro 869
1321 .....CAAGTGAATCCTATAGATGAAC 1343
870 ProArGInSerProSerArGSerSerSerProGlnProLysValLysAl 886
1344 CGTCTTATCTTAAGGTTCTGTCGATCGGCTCATCTTGTCTATTA 1393
886 allelSerProArGInArGSerHisSerLysSerSerProSerP 903
1394 CTGCGAGAAATTCATACGCAAAATTCACAGGCAAGTAGATCAGAT 1443
903 roSerArGValThrSerArGThrThrProArG.....ArGSerArG 917
1444 ATCCCACTTAAATATCTCTCTCTCAGCAGCGCTACCAAAAGCTTAA 1493
918 ValSerProCysSerAsnValGlnSerArGLeuProArGThrSerHi 934
1494 TAATGCA 1500
934 sSerGly 936

```

seq_name: sp_human:Q9UQ40

```

seq_documentation_block:
ID 09UQ40 PRELIMINARY; PRT; 1262 AA.
AC 09UQ40:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds."
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB016087; BAAB3713.1; -.
DR InterPro: IPR002965; P.Rich.extensn.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
FT NON_TER 1262
FT SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;

```

alignment_scores:

Quality:	173.50	Length:	537
Ratio:	0.700	Gaps:	27
Percent Similarity:	46.182	Percent Identity:	24.209

alignment_block:

US-09-303-518D-465 x Q9UQ40 ..

Align seg 1/1 to: Q9UQ40 from: 1 to: 1262

```

35 CCATACGTGGCAGTGTGCTCCGATGCATGACACAGCCCTCAGATTGTCGA 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 ProGlnTrp.....ArgArgSerArgSerAlaGlnArgTrp... 312
85 AACGATTTCTTTATCCGGAGSTTCTCGACCTCAGCATTTTCGAACCCGA 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 .....GlyArg.SerArgSerProGlnArgArgGlyArg 323
135 CGGGAATACCACTATTGCGACAGGGGGAGACTTCCGAGCGGCACAG 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 SerArgSerPro.....GlnArgProGlyTrpSerArgSerArg 337
185 GTCATATGCGATTGGGAACATACAAACCATCAGTTGGCAACCTGTTCC 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 mThrGlnArgArgGlyArgSerArgSerAla..... 347
235 ATCCAGACGGCGGCATTAAAGGAATATGCGCTACAT.....TGTCGG 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 .....ArgArgGlyArgSerHisSerArgSerPro 357
279 CTTTTCGATCAGGGCAGAGTCCATTCCTCCGACAAACCATCGCT 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 AlaThrArgGlyArgSerArgSerArgTrpProAlaArgArg..... 371
329 CACATTCGATTCTGATACACCGGTACTCCCGTTGACGATTCAGCCTT 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 .....GlyArgSerArgSerArgTrp 379
379 TACCGCATCATTTGGAGCGATACGAACACCATCCCGCGACGCTATGA 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379 roAlaArgArgArgSerArgSerArgTrpProAlaArgArgSerArg 395
429 CGGGCCACAGGGCGCGCTATCCCGCTCCCAAGAGCGGAGGATATAT 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
396 SerArgTrpProAlaArgArgGlyArgSerArgSerArgTrpPro...Al 411
479 ACAGCTACGACATAAAGGCGTGGCCAAATATCCGCTCAACCTGACC 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
411 AlaArgArgSerArgTrpArgSerProValArgArgSerArgSer 428
529 GACACCGCAGCAGCCGACAAACGCTTGTGACCG...TTTCCAAATAC 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 rArgSerProAlaArgArgSerArgTrpProAlaArgGlyArgSerArg 444
576 CGGTAGTATGCTGACGCAAGAGT.....AGCGACGATTCATAAC 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 ArgGlyArgSerArgSerArgTrpProAlaArgGlyArgSerArgSe 461
617 GCGCCACCCGATACAGCCCGAGCTGAGATCGGGCANTGGCCG...C 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
461 rArgTrpPro.....AlaArgArgSerGlyArgSerArgSer 474
664 GAACTTTCAACGGCAGCTAGATATGTCMAAACATCATCGCGCGCG 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474 rGlnProAlaArgArgGlyArgSerArgSerArgTrpProAlaArgGly 490
714 AGGAGAAATTTGTCG.....CGCAGCGGATCCGTCG 745
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
491 ArgSerArgSerArgSerLeuValArgArgGlyArgSerHisSerArgTr 507
746 AGGGTATACGGAAGCTCAACATTGCTGTATGACAGCGCTGGGGTTC 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 rProGlnArgArg.....GlyArgSerGlySer 517
796 CTTTCCACCGCAAAA...CAAGATGGCGCGCATACAGATTGGCAGATAT 842
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 eArgSerGlnArgGlySerArgTrpSerGlnArgSerArgSer 533
843 GCGCAACTCAAGAGCTATGCGCAGC.....AGCCATCCGCGA 881
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
534 AsnSerSerProGlnMetLysLysSerArgGlnSerSerArgTrpSerArg 550
882 .....TTGGGAGTCAAAAACCCCAATGCGGCACAAAGCATAGAACCGTTC 927

```

```

550 gSerLeuSerSerProArgSerLysAlaLysSerArg...LeuSerLeu 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
928 AGCAATATCTTTACGGCAGTCATCC..... 953
566 rArgSerLeuSerGlySerSerProCysProLysGlnLysSerGlnTr 582
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
954 .....GTCMAAGGATTTGAGCTGTTCGGGGAATATACGGCTTG 994
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 ProProArgArgSerArgSerGlySerGlnProLysAlaLys..... 597
995 GCGGCATCAGCGCATCTGTCAGCGGTCGACAGATGGCGGAGTCCGA 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 .....SerArgTrp.ProProArgArgSerArgSerSerSerPro 611
1045 TTGCGCAAGAGGAATCCGGCGCTCAGCAGCAATTTGCCGATCGGCANA 1094
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
612 ProProLysGlnLysSerLysTrpProSerArgGlnSerHisSerSe 628
1095 CGGCAATATACCGCTCCCTTACCATCCGAAATATCCGTTCAACTTGG 1144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
628 rSerProHisPro.....LysValLysSerGlyTrp 639
1145 AGCAGCGTTACGGCAAGAAACATCATCCTCTCAACCGTCGCGCGTCA 1194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
639 roProArgGlnGly.....SerLeuTrpSerProGlnAla 650
1195 AACGGAAGATGTGAATCTGCAACAGCAACAAACGCCACCGAAGCAAA 1244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
651 AsnGlnGlnSerValThr.....ProGlnArgArgSe 661
1245 GCGGTTTGAAGGTAAGGTTTCCGAATTTTGAAGAAAGCGTA..... 1287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 rCysPheGlnSerSerProAspProGlnLeuLysSerArgTrpProSera 678
1288 .....AATACGATACGAGATTAATACCGCTGTACCA 1320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
678 rGlnHisSerCysSerGlySerSerProArgValLysSerSerTrpPro 694
1321 .....CAAGTGAATCCATATGATGAAACC 1343
695 ProArgGlnSerProSerArgSerSerSerProGlnProLysValLys 1411
1344 CGTCTTTAATCCTAAGATTCTGTCGATCGGCTCATCTTGTGCTATPA 1393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
711 aLeuLeuSerProArgGlnArgSerHisSerGlySerSerProSer 728
1394 CTGCCGAAATTCATATCGCAAAATTAACAGGCAAGTGAATTCAGATAT 1443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 roSerArgValThrSerArgTrpTrpProArg.....ArgSerArgSer 742
1444 ATCCCACTAAATAATTAATCTCTCTCTCAGCAGCGCTACCAAAAAGCACT 1493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
743 ValSerProCysSerAsnValGlnSerArgLeuProArgTrpSerH 759
1494 TAATGCA 1500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
759 sSerGly 761

```

seq_name: sp_human:Q9UHA8

seq_documentation_block:

ID Q9UHA8 PRELIMINARY; PRT; 2296 AA.

AC Q9UHA8;

DT 01-MAY-2000 (TRMBLrel. 13, Created)

DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)

DE SPLICING COACTIVATOR SUBUNIT SRM300.

GN SRM300.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCBI_TaxID=9606;


```

1245 GCCGTTTCAGCGTAAAGGTTTCGAAATTTGAAAGACGTA..... 1287
      |||:::|||||:::|||||:::|||||:::|||||:::
871 rCySpHgluSerSerProAspPrpIuLeuYseSerArgThrProSera 888
1288 .....AAATAGCATAGCAGATTAATACCGCTGTACCA 1320
      ::|||:::|||||:::|||||:::|||||:::
888 rGHisSerCySerGlySerSerProProArgValIlySerSerThrPro 904
1321 .....CAAGTGATTCCTATAGATGAAAC 1343
      :::|||||:::|||||:::|||||:::
905 ProArgGluSerProSeraIySerSerSerProGlnProIyValIlyAl 921
1344 CGNCTTAAATCCTAAGAGTATCTGTGCAAGCGCCTCATCTCGTCTATA 1393
      :::|||||:::|||||:::|||||:::|||||:::
921 aIleIleSerProArgGlnArgSerHisSerIySerSerProSerP 938
1394 CTGCCAGAAATTCATATCGCAAAATTAACCAAGCGCAAGTAGATCAATAT 1443
      ::|:::|||||:::|||||:::|||||:::
938 roSerArgAlaThrSerArgThrThrProArg.....ArgSerIySer 952
1444 ATCCCACTTAAATATTCATCTCTCTTAGAGACGCGTACCAAAAGGACATA 1493
      ::|||:::|||||:::|||||:::|||||:::
953 ValSerProCySerSerAsnValGluSerArgLeuLeuProArgTyrSerHi 969
1494 TAATGGA 1500
      :::|||||
969 sSerGly 971

```

seq_name: sp_fungi:Q9P6T1

```
seq_documentation_block:
ID      Q9P6T1      PRELIMINARY;      PRT;      1992 AA
```

DT 01-DEC-2000 (TREMblrel_15, Created)
 DT 01-DEC-2001 (TREMblrel_19, last sequence update)
 DT 01-DEC-2001 (TREMblrel_19, last annotation update)
 DE HYPOTHETICAL 207.4 KDA PROTEIN.
 GN 1566.220.
 OS *Neurospora crassa*.
 CC *Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;*
 CC *Sordariales; Sordariaceae; Neurospora.*
 CX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Algin V., Hehseisel J., Brandt P., Fartmann B., Holland R.
 RA Nyakatura G., Mewes H.W., Mannhaupt G.,
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German *Neurospora* genome project;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AL353822; CAB88653.2; -
 KW Hypothetical protein.
 SQ SEQUENCE 1992 AA; 207432 MW; CAb8145FD154281B CR64;

alignment_scores:		
Quality:	170.00	514
Ratio:	0.742	24
Percent Similarity:	44.553	Percent Identity: 23.735

```
alignment_block:
US-09-303-518D-465 x Q9P6T1 . .
```

Align seg 1/1 to: Q9P6T1 from: 1 to: 1992

```

125 TCGAACCCGACGGGAATACCACTTCGCGAGACGGGGGAATTTGGC 174
      :::::||||:::  :::  :::::||||  |||  |||
1261 AATATPCTAATLAPROVALHISATLAFTHRALTHRPROGLYALALeupr 1277HVALeupr
      :::::||||:::  :::  :::::||||  |||  |||
175 GAGCCGACGGGTCTATCGGATTGGGAACATACAAAGCCATCAGTGGG 224
      |:::  :::::  :::::  :::::  :::::  :::::  |

```

[illegible]


```

483 CTACGACATATAAGCGC.....TTCGCCAATAATCCCGCTCAACC 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3377 hTtrpThrValProAlaGlnThrThrThrProMetSerThrMetSerThr 3393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
524 TGACGACACACCGCAGCAGCGAGC...AACGGCTTGCGACCGCTTCCAC 570
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3394 ILeHstHrSerSerThrProGlnThrHstHrSerThrValLeuThr 3410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
571 AATPACCGAGTAGTGTGACGCAAGAGTAGCGCAGCGATTCACACGCGC 620
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3410 ThrThrAlaThrMetThrArg.....AlaThrAsnSerThrAlaAT 3424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
621 CACCGATACAGCCCCGAGCTGACAGATCGGCAATCGCGCGCAAGCTT 670
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3424 hTProSerSerThrLeuGlyThrThr.....ArgTlle 3434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
671 TCAAGGCACTGCAGATATCCGTCAAAACATCAGCGCGCGCAGAGAA 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3435 LeuThrGluLeuThrThrThrAlaThrThrAlaAlaThrGlySerThr 3451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 ATTGTGCGCGCAGCGCATCCGTGCGAGGGTATTAAGCGAGCTCAACAT 770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3451 rAlaThrLeuSerSerThrPro..... 3458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
771 TGCTGTATGCAAGCGCTGGGCTTGTCTTCCACCGCAAAACAAGATGGCGC 820
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3459 .....GlyThrThrThrPileLeuThrGluPro..... 3467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
821 GCATCAGCATTTGGCAGATATGGCGCACTCAACAAAGACTATGCCGACGA 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3468 ..SerThrIleAlaThrValMetValProThrGlySerThrAlaThrTh 3483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
871 GCCATCCGCGATGGCGAGTCCCAAAACCCCAATGCCGACAAGCATAGA 920
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3483 rSerSerThrLeuGlyThrAlaHstThr..... 3492
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
921 AGCGGTGACGATATCTTACGCGACATCATCCCGTCAAAAGATGGAG 970
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3493 ..ProLysValValThrAlaMetAlaThrMetPro..... 3503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
971 CTGTTCGGGGAATAATPACGGCTTGGCGCATCAGGCGCATCCTGTCAAG 1020
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3504 .....ThrAlaThrAlaSerThr..... 3509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1021 CGGTGCGAGATGGCGAGATCGCATGCCGAAGAAATCCGCGCTAG 1070
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3510 .....ValProSers 3513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1071 CGACAAATTTGCGATGGCGCATCGCAATAACCCGCTCCCTTACCAT 1120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3513 erSerThrValGlyThrThrArgThrProAlaValLeuProSerSerLeu 3529
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1121 CCCGAAATATCCGTCAACCTTGAGCAGC.....GTTACGGGCAAA 1161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3530 ProThrPheSerValSerThrValSerSerSerValLeuThrThrLeuAr 3546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1162 GAATACATCACTCCCTCA.....CCGTGC 1186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3546 gProThrGlyPheProSerSerHstPheSerThrProCys 3559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: sp_rident:P70670

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66312450; PubMed=8698236;
RA Yotov W.V., St-Arnaud R.;
RT "Differential splicing-in of a proline-rich exon converts alphaNAC
RT into a muscle-specific transcription factor.";
RL Genes Dev. 10:1763-1772(1996).
DR EMBL; U48364; AAB18734.1; -.
DR EMBL; U48363; AAB18732.1; -.
DR MGD; MGI:106095; Nacc.
DR InterPro; IPR002715; NAC.
DR InterPro; IPR000449; NAC.
DR Pfam; PF01849; NAC; 1.
DR Pfam; PF00627; UBA; 1.
SQ SEQUENCE 2187 AA; 220599 MW; 003646AA864DEBFD CRC64;

```

```

alignment_scores:
    quality: 156.50      length: 567
    ratio: 0.660         gaps: 24
    percent similarity: 41.799      percent identity: 20.988

```

alignment block:

US-09-303-518d-465 x P70670 ..

Align seg 1/1 to: P70670 from: 1 to: 2187

```

167 AACTGGCGGAGCGCGATGATGATGGGAACATACCAACCAT 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 SerIleProLysValThrSerProSerProGlnLysThrProLysSerVa 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 CAGT.....TGGGCAACCTGTTTCATCCAGCAGCGCCCATTAAGAA 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
906 IserLeuLysGlyAlaProAlaMetThrSerLysLys...AlaThrGluI 922
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
261 TATCGGCTACATGTGCCGCTTCCGATCAGCGGCGCAGATTCATGCC 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 LeuAlaSerLysAspValSerProSerGlnPheProLysGluValPro 938
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 CCTTCGCAACACATGCCATCCGATTCGATTCGATGAACCGGTAGTCC 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
939 .....LeuLeuGlnHstValProProThrSerProProLysSerPr 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
361 GTTGACGAGATTCA.....GCCTTACCGCA..... 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
952 oValSerAspThrLeuSerGlyAlaLeuThrSerProProProLysGlyP 969
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 .....TCCATTTGGAGGATACGACACCATCCG.....CCGAGC 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
969 roProAlaThrLeuAlaGlnThrProThrTyProLysLysSerProLys 985
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 GCTATGACGGGCGACAGGGCGGCGCTATCCGCTCCGAAGCGCGAGG 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
986 ProAlaAlaSerLysLysThrProAlaThrProSerProGluGly.... 1000
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
472 GATATATACGCTACGACATATAAAGCGCTGCCCAAAATATCCGCTCAA 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 ....ValThrAlaValPro...LeuGluIleProProCysSerLysLys 1015
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
522 CCTGACCGCAACCGCAGACCGGAGCGGTGTGCGCGTTTCCACA 571
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1015 Ia...ProLysThrAlaAlaProLysGluSerSerAlaThrSerSer... 1029
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
572 AATCCGAGTAGTATGCTGACGCAAGAGTAGCGACGATTCACACGCGCC 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1030 ..... 1033
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
622 ACCGATACAGCC.....CCGAGCTGGACATCGGCGCA 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1033 aProLysThrAlaValSerLysGluIleProSerLysGlyValThrAlav 1050
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

657	TCGCCCGCAGCTTTCAACGGCAGCTGCAGATATCGTCAAAACATCATG	706
1050	ATProLeuGlnLeuSerLeuProLeuysgluThrSerIysSerAlaThr	1066
707	GGCGGCGAGGAAATGTCGGCCAGCGCATGCCGTGACGGGTATTAAC	756
1067	ProGlyIuThrSerAlaSerProIysSerProIysThrAlaGln	1083
757	GAAAGCTCAACA.....TTCGCTTATGACGGCTTGAGTCT	794
1083	YProIysgluThrProProGlyIglValThrAlaValProProGluIle	1100
795	GCTTTCACCCGAAAACAGATAGCGCGCATCAACGATTGGACGATATG	844
1100	erLeuProProIysgluThrProGlnsmAlaThrProAsnGluSerLeu	1116
845CGCACTCAAGACT.....	859
1117	AlaAlaSerSerGlnIysArgSerProIysThrSerValProIysgluThr	1133
860ATGCCGACGACCAT	875
1133	rProProGlyIglValThrAlaMetProLeuGlnIleProSerAlaProG	1150
876	CCGCGATTTGGGAGTCCAAACCCA.....	901
1150	IuIysAlaProIysThrAlaValProIysGlnIleProThrProGluAsp	1166
902ATGCCGACGACCATAGAA	921
1167	AlaValThrIleLeuAlaGlySerProLeuSerProIysAlaSerIly	1183
922	GCCGTCAGCAATATCTTTACGGCAGCTATCCCGCTCAAAAGGATTGGAG	971
1183	sThrAlaIaIaProIysgluAlaIaProAlaThrProSerValGlyAla....	1198
972	TGTTCCGGGAAATACGCTTGCGCGCATCACGGCATCTCCTACAGC	1021
1199IleAlaValSerGlyIuIleSerProSer	1208
1022	GCTCGAGATGGGCGCATCGCATTCGCCGACAAAGGAAATCCGCGTCAGC	1071
1209	ProIysThrSerIysThrAlaIaIaProIysgluAsn.....SerAl	1223
1072	GACAATTTTGGCGATCGGCATACGCCAAATACCCGCTCCCTTACATCT	1121
1223	atThrLeuProProIysArgSerProIysThrAlaIaIaProIysgluThrP	1240
1122	CCGAAATATCCGTTCAACTGTGAGAGCGAGGTACGGCAAG...AAACA	1168
1240	roAlaThrSerSerGluGlyValThrAlaValProSerGluIleSerPro	1266
1169	TCACCTCTTCACCGTCCG.....	1188
1257	SerProProThrProIaSerIysgluAlaProValThrLeuThrProIy	1273
1188	1188
1273	sgIyAlaProAsnAlaLeuAlaGluSerProAlaSerProIysIysValP	1290
1189CCGTCAACGCGAAG	1203
1290	roIysThrAlaIaIaProGluGluThrSerThrProSerProGlnIys	1306
1204	AATGTGAACCTGGCA.....AACAA	1223
1307	IleProIysValAlaGlyProIysgluAlaSerAlaThrProProSerIy	1323
1224	ACGGCACCCGACCAACCAATGCGCTTGACGGTAAAGGTTTCCGAAT	1273
1323	sluYThrProIysThrAlaValPro.....	1331

```

1274 TTGAAGAGAGCTAAATAATACGATACGAGAAATTAATACCGGTGACCA... 1320
      |||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
1332 .....LysGIunThrSerAlaProSerGIunGIyAlaThrAlaValAlaProLeu 1346
1321 .....:::CAAGTGAATCCTATAGA 1337
1347 GIUnIleProProSerProArgLysAlaProLysThrAlaAlaProLysGI 1353
1338 TGAAGCCGTCCTTAAATCCATAAGGCTTCGTGCGAGTCGCTCATTCCTGGCT 1387
      |||:::||||:||||:||||:||||:||||:||||:||||:||||:
1363 uThrProAlaProSerProGIunGIyAlaThr..... 1373
1388 CATAAAGTCCACGAATTCAATCGCAAAATTTCCAAAGCAGGTAGATC 1437
      ||||||:::|||||:::|||||
1374 .....ThrAlaProValGIunIleProProSerProArgLysGIySerLys 1388
1438 AGATATATCCACACATTAATAATTACCTCTCCTTAGCAGCGCTACCAAAAGG 1487
      ::|:::||||:||||:||||:||||:||||:||||:||||:
1389 Lys.....AlaGIySerLysGIunThrProIlnThrProSerProGIunGI 1403
1488 A 1488
1403 Y 1403

```

seq_name: sp_mammal:Q29071

```
seq_documentation_block:
ID Q29071 PRELIMINARY; PRT; 528 AA.
```

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GASTRIC MUCIN (FRAGMENT).
OS Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
NCBI_taxid=9823;
NCBI_taxid=9823;

RP SEQUENCE FROM N.A.
RC STAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA.
RX MEDLINE-95275264; Pubmed 7455593;
RA Turner B.S., Bhaskar K.R., Hadzopolou-Cladaras M., Specian R.D.,
RT Lakont J.T.;
RL "Isolation and characterization of cDNA clones encoding pig gastric
mucin.";
Biochem. J. 308:89-96(1995).

RP SEQUENCE FROM N.A.
RC STAIN-XYOKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS: TISSUE=GASTRIC MUCOSA
RX MEDLINE=94102478; Pubmed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopolou-Cladaras M., Spectan R.D.,
RA Lakont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat.";
RL Gastroenterology 106:200-200(1994).
DR EMBL: U10281; AAC48526.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 528 AA; 49907 MW; 0BF0F6879203B2EA CRC64;

```
alignment_scores:
  Quality: 151.00
  Ratio: 0.599
  Percent Similarity: 51.220
  Length: 4922
  Gaps: 22
  Percent Identity: 23.5777
```

alignment_block:

Align seg 1/1 to: Q29071 from: 1 to: 528

119 AGCATTTCGAACCCGACGGGAATACCACC...TATTCGGCAGCAGG 162
||| |||::||| |||::||| |||::||| |||::|||

```

9 SerSerSerSerProthThrSerThrThrSerValGlnSerSerSer 25
163 GGGGAACTTCCGAGCGCGATCATCG..... 193
25 rSerSerValProthProSerThrThrSerValGlnProSerSer 42
194 ..GATTGGAAACATACAAAGCCATCATGGGCAACCTGTTCAATCAGC 241
42 lYSerAlaProthThrSerAlaThrSerValGlnThrSerSerSer 58
242 AGGGCGCCATTAAAGAAATATCGCTACATTG..... 274
59 SerProPro.....lSerSerThrIleSerValGlnThrSer 72
275 ...TCGGCTTTCCGATCAGCGGCGAGATCGATCCCTTGAGCA 320
72 rSerSerSerValProthThrSerThrThrSerValGlnProSer 89
321 CCATGCTTCACATCCGATTGTGATGAAGCCGGTAGTCCCGTTGACG... 367
89 eSer.....SerAlaProthThrArgAlaThrSer 99
368 .....GATTACGCTTACCCCATCATCGAGCGATAGCAACAC 408
100 ValGlnSerSerSerSerSerAlaProthSerSerThrThrSerVa 116
409 CATCCCGCGGAGCGCTATGACGGCGGCGGCTATCCCGCTCC 458
116 lGlnProSer.....SerIleSerValProthRT 127
459 CAAGCGCGGAGGATATATACAGCTACACATAAAGCGTTGCCCAA 508
127 hSerAlaThrSerValGlnSerSerSerSerSerAlaProthThr 143
509 ATATCGGCTCAACCTGACCGACAGCCGACGCGGCAACGGCTGTC 558
144 SerAlaThrSerValGlnProSerSerSerSerProthIleSer 160
559 GACCGTTTCCAAATACCGGTAGTGTGACGCAAGAGTAGGCGGCG 608
160 rThrValSerValGlnPro.....SerSerS 169
609 ATTCAAGCGCGCACCCGATACAGCCCGAGCTGG.....ACAGATCG 652
169 eSerSerAlaProthThrSerAlaThrSerValGlnProSerSer 185
653 GCATGCGCGGCAAGCTTTCAGC...GCACTGCAGATATCGTCAAA.. 697
186 SerSerProthIleSerSerThrValSerValGlnThrSerSer 202
698 .....ACATCATCGGCGGCGGAGAGAAATGTGCGCGCGAG 734
202 rSerValProthThrThrSerThrThrSerValGlnProSerSer 219
735 CGATGCGG.....TGCAAGGTATAAGGAGAGGATCAAA 766
219 eValProthThrThrSerAlaThrSerValArgSerSerSerSer 235
767 ACATGCTGTATGACAGCGCTGGGTC.....TCCTTTC 801
236 ThrProthProSerThrThrSerValGlnProSerSerSerSer 252
802 ACCGAAACAAAGATGGCGGATCAAGATTGGCAGATATGGCGCACT 851
252 aProthThr.....SerAlaThrSerValGlnProSerSerSer 267
852 CAAGACTATGCGGAG.....CAGCATCCGATTGGGCGAG 889
267 eThrProthProSerThrThrSerValGlnProSerSerSerSer 283
890 TCCGAAACCCCA.....ATCGCGCACAGGACATAGAACCC 924
284 AlaProthThrSerAlaThrSerValGlnProSerSerSerSer 300

```

```

925 G.....TCACAAATATCTTACGGCAGTCATCCCGTCAAGAGAT 965
1 oProthSerSerThrIleSerValGlnProSerSerSer..... 314
966 TGGAGCTGTGGGAAATATAGCGCTTGGCGGATACGGCACATCTG 1015
315 .....SerSerProthThrSerThrThrSerValGlnProSer 327
1016 TCAAGCGGTGCGACATGGCGAGATCG.....CATTGCGC 1050
328 SerSerGlySerAlaProthThrThrSerAlaThrSerValGlnPro 344
1051 AAAGGAATATCCCGCTGACGACAAATTTGCCAGTGGCCATAGCCAA 1100
344 rSerSerSerProthIleSerSerThrIleSerValGlnProSer 361
1101 ATACCGTCCCTTACCATTCGCCGAATATCCGTTCAACTTGAGCAGC 1150
361 eSerSerSerProthThrThrSerThrThrSerValGlnProSer 377
1151 GTTACGGCAAAAGAAACATCAC.....TCCTC 1178
378 GlySerAlaProthThrThrSerAlaThrSerValGlnProSerSer 394
1179 AACCGTCCCGCTGCAACGCAAGATGTGAAGTGGCAACAAAGCC 1228
394 rSerValProthThrThrSerAlaThrSerValArgSerSerSer... 409
1229 ACCGAGACCAAGTGCCTTGTGACGTTAAGGTTTCGATTTTGA 1278
410 ..SerSerThrProthProSerThrThrThrSerValGlnProSer 425
1279 AAAGACGTAAATATACAT.....ACGAGATTAATACGCTGTACACA 1322
426 SerSerValProthThrThrSerAlaThrSerValGlnThrSerSer 442
1323 AGTGAATCTATATAGATGACACCGCTTTTAATCTTAAGTCTGTGAT 1372
442 rSerThrProthProSerThrThrSerValGlnProSerSerSer 459
1373 CGGCTATCTTGTCTATATCTGCC 1398
459 eAlaProthThrThrSerAlaThrSer 467

seq_name: sp_bacteria:p94909
seq_documentation_block:
ID P94909 PRELIMINARY; PRT; 529 AA.
AC P94909;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 58.7 KDA PROTEIN (FRAGMENT).
OS Microbacterium ammoniaphilum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococineae; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=42460;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15354;
RX MEDLINE=96257250; PubMed=8654988;
RA Striebel H.M., Seeder S., Jarsch M., Kessler C.;
RT "Cloning and characterization of the Mami restriction-modification
RL system from Microbacterium ammoniaphilum in Escherichia coli.";
DR EMBL, X79027; CAA55649.1; -
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 529 AA; 58716 MW; 5DC1216237DF772D CRC64;

```

alignment_scores:

Quality: 153.50 Length: 427
Ratio: 0.872 Gaps: 22
Percent Similarity: 41.218 Percent Identity: 25.293

Alignment block:

US-09-303-518D-465 x P94909 ..

Align seg 1/1 to: P94909 from: 1 to: 529

```

100 CGGAGAGTT.....CTGACCGCTCATTTGCAACCCGAGGGAATA 143
|||||
78 ArgGlnValHisProLeuGlnArgProHisGlnIupro..... 90
144 CCACTATTCGGCGAGGAGGGAATTGC.....CGAGCGC 180
|||||
91 .....GlyAlaArgGlyGlyIuproValArgAspAspArgIaG 104
181 AGCGGTATATCGG.....AFTGGGAACATACCAAGCA 215
|||||
104 IuaArgGlyArgGlyIuproAlaArgProAlaAlaArgAspAlaArg 120
216 TCAGTTGGGCAACCTGTTCATCCAGCA...GGCGCCATTAAAGGAATA 262
|||
121 HisLeuArgLeuGlyAlaHisProAlaArgGlyValLeuArgArgHis 137
|||
263 TCAGCTACATTCGCGCTTTCCGATCAGGGCA...CGAATCCATTC 309
|||||
137 sarArgHis.....ArgAlaArgGlyIuaArgGlyArgGly 150
|||||
310 CCGTTGCAACCAATGCGCTCATTCGATTCGATGAGCGGTAATCC 359
|||||
150 roArgGlnGln..... 153
360 CGTTGACGAGATTACGCTTTACCCGATTCATGGAGCGATACGACACC 409
|||||
154 .....ValProArgGlnHisProArgGlyArgArgAspArgAl 166
|||||
410 ATCCGCGGAGCGCTATGACGGGCGACAGGGCGCTATCCGCGTCC 459
|||||
166 aGlyArgProGlyLeu.....HisArgArgArgArgArgAla 179
|||||
460 AAAGCGCGAGGAGATATACAGCTACAGACATAAAGCGCTGCCCAAA 509
|||||
179 rGArgArgGlnGlyArgProGlnValArgHis..... 189
510 TATCCGCGCTCAACCTGACCGCAACCGGACGACCGGACGCGTTGCG 559
|||||
190 .....GlyAspArgGlnHisArgAlaAspProArgArg 200
560 ACCGTTTCACAATACGCGTAGTAGTCGACGCAAGAGTAGGCGACGA 609
|||||
200 gProArgAspProArgAlaGlyHisHisHisProLeuArgGlnGlyGlnG 217
610 TTCACAACCGCGACCGGATACAGCCCGAGCTGACAGATCGG..... 653
|||||
217 LysGlnGlnAlaArgProLeuArgAlaArgGlyGlnGlyGlyProGly 233
653 ..... 653
234 GlyProAlaGlyArgHisProAlaLeuGlyAspGlnAspArgProArgAl 250
654 .....CAATGCGCGCGAGGCTTTCACAGGCGACGATATGCTCAAAAC 699
267 IuaArgArgArgGlyArgAlaAspGlyCysArgAlaGlnGlyArgAlaArg 283
732 AGCGGATCCGTCAGGCTATAGCGGAAGGCTCAACACTGCTGTATAGC 781
284 ArgAlaArgArgAlaGly..... 289

```

```

782 ACGGCTTGGGTCTGCTTCAACCGAAACAGAT.....GGCGCGAT. 824
|||||
290 .....GlyGlyGlyLeuProAlaArgGlnAspArgValArgAlaHisA 304
825 .....CAACGATTTGGCAGATATGCGGCACTCAAAAGACTATGCCGACG 869
|||||
304 rGProArgProArgGlyArgGlnGlyAlaAla...ArgLeuAspArgAla 319
870 AGCCATCCGCGATTGGGCGAGTCCAAACCCCAATGCCGACAGGATAG 919
|||||
320 GlyGlyValGlyAlaArgProAlaArgProArg...ArgLeuArg..... 333
920 AAGCGGTGACGAATATCTTTACGCGAGTCATCCCGCTCAAGAGGATTTGA 969
334 .....HisProArgProAlaAspLeu 341
970 GCTGTTCGGGGAATAATACGCGTTGGGCGG...CATCAGCGCATCTCTGT 1016
|||||
341 roHisGlyGlyProGlnGlyValAlaArgLeuAspHisProGlnGlyLeu 357
1017 CAAGCGGTGCGAGATGGCGGAGATGCGCATTTGCCGAAGGAATCCGCG 1066
358 GlnGlyAlaAlaGlySerArg..... 364
1067 TCAGCGCAATTTTGCAGTCGCGCATACGCCCAATACCCGCTTAC 1116
364 ..... 364
1117 CATCCCGAATATCCGTTCAAACTTGACAGCGCTTACGCAAGAAATA 1166
|||||
365 .....ArgHisProHisArgArgAlaArgLeuHisGlnGlyArg 378
1167 CATCACTCTCTCAACGCGTCCGCGCTCAACGAAAGAAATGTGAATCTG 1216
379 GlyAspLeuLeuArgArgProArgArgAspArg.....LeuG 391
1217 CAACCAACCGCACCCGGAAGACCAAGTCC 1247
391 yArgArgGlyProArgGlnGlyGlnGlyAla 401

```

seq_name: sp_human:Q14879

```

seq_documentation_block:
ID Q14879 PRELIMINARY; PRT; 543 AA.
AC Q14879;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE MUCIN (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEOBRONCHIAL; MUCOSA;
RX MEDLINE=9343858; PubMed=7916618;
RA Dufosse J., Porchet N., Audie J.P., Guyonnet Duperrat V., Laine A.,
RA Van-Seuningen I., Marikich S., Degand P., Aubert J.P.;
RT "Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic
RL Biochem. J. 293:329-337(1993).
DR EMBL; X74370; CAA52408.1; -.
FT NON_TER 1
FT NON_TER 543
SQ SEQUENCE 543 AA; 55197 MW; 6767A5E3518B287B CRC64;

```

alignment_scores:
Quality: 149.50 Length: 395
Ratio: 0.808 Gaps: 16
Percent Similarity: 46.835 Percent Identity: 22.025

alignment_block:

US-09-303-518D-465 x Q14879 ..

Align seg 1/1 to: Q14879 from: 1 to: 543

```

128 AACCCGAGGGGAAATACCACTATTGGCAGAGGGGAACTTGGCCGAG 177
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
44 SerProLeuAspSerThrAlaThrProSerThrProGlyThrAlaPro 60
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
178 CGCAGCGGTCATTCGAGTTGGGAAACATACAAAGCCATCGAGTGGCAA 227
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
60 oProLysValLeuThrSerProAlaThrProThrAlaThrSerSerL 77
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
228 CCTGTTCATCCAGCAGCGGCCCATTAAGAAATATCGCTACATTTGCC 277
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
77 yAlaAlaThrSerSerSerPro.....ArgThrAlaThrThrLeuPro 91
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
278 GCTTTCCGATCAGCGGCGAGCAAGTCATTCGCCCTTCGACACCATGCC 327
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
92 ValLeuThrSerThrAlaThrLysSerThrAlaThrSerValThr.... 106
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
328 TCACATTCGCATTCGATGAAGCCGGTGTCCCTTCAGCGATTACGCT 377
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
107 .....ProLeuProSerSerThrLeuGlyThr 115
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
378 TTACCGCATCCATTGGAGCGATACGAACCATTCGCCCGCGAGGTATG 427
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
115 hTrThrGlyThrLeuProGluThrThrThrProValAlaThrMetSer 131
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
428 ACGGCCACAGGCGCGCGGTATCCCGTCCCAAGCGCGAGGATATA 477
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
132 ThrLeuHisProSerSerThr.....ProGluThrThrHisThrSe 145
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
478 TACAGCTCAGACATAAAGGGCTTGCCCAAAATATCCGCCCAACCTGAC 527
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
145 rThValLeuThrThrLysAla..... 152
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
528 CGACAAACCGCAGCAGCAGCAACGCGCTTTCGACCGTTCCCAATACG 577
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
153 ..ThrThrArgAlaThrSerSerThrSerThrProSerSerThrPro 168
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
578 ...GTAGATCTGACCGCAAGAGTATAGCGAGTTCAAACGCGCCAGC 624
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
169 GlyThrThrThrPheLeuThrGluLeuThrThrAlaAlaThrThrAla 185
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
625 CGATACAGCCCGCAGCTGAGCAGATCGGCAATCGCGCCGAG.....C 668
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
185 aglYThrGlyProThrAlaThrProSerSerThrProGlyThrThrTri 202
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
669 TTTCACAGCGCAGCTGCAGATATCGTCAAAACATCATCGCGCGCAGAG 718
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
202 leuThrThrGluLeuThrThrThrAlaThrThrAlaSerThrGlySer 218
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
719 AAATGTGCGCGCAGCGAGTCCGTGAGGGTATAGCGAAGCTCAAC 768
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
219 ThrAlaThrLeuSerSerThrPro..... 226
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
769 ATTGCTGTATGACAGCGCTTGGCTGCTTTCACCGCAAAACAAGATGC 818
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
227 .....GlyThrThrThrPheLeuThrGluProSerThr..... 237
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
819 GCGCATCAACGATTTGGCAGATATGCGCACTCAAGACTATGCGCAG 868
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
238 ..ThrAlaThrVal.....ThrValProThr 245
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
869 CAGCATCCGCGAGTTGG.....CAGTCCAAACCCCATTCGCCGA 909
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
246 GlySerThrAlaThrAlaSerSerThrGlnAlaThrAlaGlyThrProH 262
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
910 CAAGCATCAAGCCGTGACGCAATATCTTACGCGAGTCCCGC.... 955
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
262 sValSerThrThrAlaThrThrProThrValThrSerSerLysProLeuL 279
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

956 ....TCAAGGATTTGGACCTGTGCGGCAAAATACGCTTGGCGCGCA 1000
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
279 euLeuSerProGlyLeuGlnLeuProSerAlaLeuArgSerThrAlaThr 295
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1001 TCAGCGCAG.....ATCCTGTCAAGCGGTGG 1026
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
296 AsnThrHisSerThrGlnLeuThrSerHisProLeuLeuProGlyThr 312
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1027 CAGATGGCGCAGATCGCATTCGCCGAAAGGAAATCCGCCGTGACGACAA 1076
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
312 rThrTrpThrArg.....LeuSerGlnThrThrThrProThrAlaThrM 327
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1077 TTTTGGCGATGCGGCATACGCCCAATACCGCTCCCTTACCATTCGCCAA 1126
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
327 et.....SerThrAlaThrProSerSerThrProGluThrValHisThr 341
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1127 ATATCCGTTCAAACTTGAGCAGCGTTCAGCGCAAAAGAAACATCA.... 1171
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
342 SerThrValLeuThrAlaThrAlaThrThrThrGlyAlaProAlaLeuThr 358
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1172 ....CCTCTCAACCGTCCCGCGTCAACGGAAGAAATGTGAACCTGCG 1217
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
358 pProProProProProGluGln..... 367
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1218 AAACAAAGCCGACCGCAGACCAACCAAGTCCGTTG 1252
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
368 .....LeuThrLeuProLysCysArgLeu 375
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

seq_name: sp_invertebrate:Q20908

seq_documentation_block:

```

ID Q20908 PRELIMINARY; PRT; 770 AA.
AC Q20908;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F56H9.1 PROTEIN.
GN F56H9.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiidae;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 274473; CA98949.1; -.
DR InterPro; IPR004024; DUF225.
DR Pfam; PF02795; DUF225.2
SQ SEQUENCE 770 AA; 81625 MW; 52FD5F6628D07752 CRC64;

```

alignment_scores:

```

Quality: 149.00 Length: 219
Ratio: 1.263 Gaps: 9
Percent Similarity: 53.881 Percent Identity: 25.571

```

alignment_block:

US-09-303-518D-465 x Q20908 ..

Align seg 1/1 to: Q20908 from: 1 to: 770

```

86 AGCATCTTTATCCGGGAGGTTCGACCGTCAACATTTGAAACCCGAC 135
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
37 ThrValProValSerThrThrThrSerThrSerThrThrThrThrTh 53
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```



```

774 erAlaProSerThrSerValValThrArgArgProThrSerThrThra 790
494 AAG..... 496
791 GlnMetSerSpGluGluArgSerArgIleAlaMetAspIleSerSe 807
497 .GCGTTCGCCAATAATCCGCC.....TCACCTGACCGACAACG 536
807 rAlaLeuProAlaProSerAlaLeuLeuTyrGlySerAsnSerThrSers 824
537 CAGACCGGACAAACGCTTGTCCAGCTTCCACAATACCGTAGATGC 586
824 erLeuProSerAlaAlaValSerThrAlaSerValPro..... 837
587 TGACGCAAGAGTAGCGCATTCAAACGCCCAACCGCATACACCCC 636
838 .....SerThrAlaAlaGAspAsnProValGluThrArgPr 849
637 GACCTGACAGATCGGGCAATGCCGCCGAAGCTTTCACGCGACTGCAGA 886
849 oSerGlnProHisValThrMetProProGlySerSerSerGlnProIle 866
687 TATCGTCA.....AAACATCATCGCGCGCGCAGAG 718
866 euSerSerGluValLeuGlnProThrArgLeuProSerAlaThrThrSer 882
719 AATGTGCGCGCGCGAGATGCCGAGGTATAGCGAGGCTCAAC 768
883 GluAlaLeuProValThrGlnProIleArgHisProSerProProValAl 899
769 ATGCGCTGATGACGCGCTTGGCTGCTTCCACCGAAACAGATGCG 818
899 arThrValIleProThrAlaValAspLysLysProValSerGlnAsnG 916
819 GCGCATCAACGATT.....TGCGAGATATGGCGCACTCAAGACTAG 862
916 lngLysSerAsnValProLeuPheAsnIleThrAsnSerSerSngLys 932
863 CCGCAGCAGCCATCCGCGATTGGGCGAGTCCAAACCCCAATCCGCA 912
933 ProGln.....LeuAsnGlyTyrProAspGly 941
913 GCGATGAGCGCGTCAAGATATCTTTGCGAGATCATCCCGTCAACG 962
941 rGlyAsnLysPheGlnAlaTyrGlyTyrGlyMetAsnTyrHisGln 957
963 GATTGGAGCTGTCGGGGA.....AAATACGCGTTGGCGGCGCATCAG 1006
957 yTyrProGlyTyrGlnGlyTyrAsnSerTyrGlyAsnGly..... 970
1007 CACATTCCTCAAGCGGTCGAGATGGCGAGATCGCATTCGCCAAGGG 1056
971 .....MetGlyGlnLeuAlaLeuThrHisAsn 979
1057 AATCCGCGCGTACGAGACAAATTTGCCGATGGCGCATGCCCAATACC 1106
980 .....AlaValThrSerLeuPr 985
1107 GTCCCTTACCATTCGCCAATATCCGTTCAAACTTGAGCAGCGTTACG 1156
985 oProLeuValProSerGluAsnArgPheSerGlyThrAlaGlnProLeu 1002
1157 GCAAGAAACATC..... 1170
1002 yGluSerAspIleMetGluPheLeuGlyThrGlnGlnArgGlnAlaGly 1018
1171 ..ACCTCCTCAACCGTCCGCGCTCAACAGCAAG..... 1203
1019 SerSerSerArgAlaValProProAlaSerAlaSerThrSerAlaAla 1035
1204 .....AATGTGAACCTGGCAACAAACGCCCGCAGACCAAG 1243

```

```

1035 rGlyIleThrAspLeuSerMetAlaAlaSpLysMetGluValLeuTyrArg. 1051
1244 TCCCGTTTGACGGTAAGGGTTTCGGAATTTGAAAAAGACGTAAATAC 1293
1052 .....GluAlaAspPheThrHis 1057
1294 GATACGAGATTAATACCGCTGTACCAAGTGAAT 1329
1058 LysGlyAsnCysAspThrMetValSerGlnCysAsn 1069
seq_name: sp_invertebrate:Q9N3Y8

```

seq_documentation_block:

```

ID Q9N3Y8 PRELIMINARY; PRT; 2344 AA.
AC Q9N3Y8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Y40C5A.3 PROTEIN.
GN Y40C5A.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Kalicki J., Randall J.;
RT "The sequence of C. elegans cosmid Y40C5A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024772; AAF60538.1;
DR InterPro; IPR000194; ATPase_alpha_beta.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR SEQUENCE 2344 AA; 245902 MW; 7D8696D48D0ADDBB CRC64;

```

alignment_scores:

Quality	Ratio	Length	Gaps
145.50	0.500	625	31
Percent Similarity: 46.560	Percent Identity: 21.600		

alignment_block:

US-09-303-518D-465 x Q9N3Y8 ..

```

Align seg 1/1 to: Q9N3Y8 from: 1 to: 2344
12 CCGCAAAATATCCCTTATTTCTGTCATACATGCGAGTGTGCTGCCGATGC 61
||| .....||| ||||| .....||| .....||| .....|||
681 ProThrThrValProLysSerThrIleThrGlyArgMetProSerSerVa 697
62 AVGCACACGCGCTCAGATTGGCAACAGATCTTTATCCGCGAGCTTTC 111
: .....||| .....||| .....||| .....||| .....|||
697 I.AsnProSerAspAsnLeuSerAsnAsnGlnMetLaspGlyAsnAlaHis 713
112 GACCGT..... 117
714 AsnGlnProSerGluPheSerAsnLeuSerGluThrSerIleThrProAl 730

```


DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE SER/ARG-RICH PRE-MRNA SPLICING FACTOR SR-A1.
 OS SR-A1.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Scottias A., Kyriakopoulou L., Katsaros D., Diamandis E.P.;
 RT "Cloning of a gene (SR-A1) encoding a new member of the human Ser/Arg-
 rich family of pre-mRNA splicing factors."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF254411; AAF87552.1;
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENSN.
 SO SEQUENCE 1312 AA; 139269 MW; 0CB1C87C963C52BD CRC64;

alignment_scores:

Quality: 145.00 Length: 543
 Ratio: 0.604 Gaps: 30
 Percent Similarity: 44.199 Percent Identity: 24.125

alignment_block:

US-09-303-518D-465 x Q9NR59 ..

Align seg 1/1 to: Q9NR59 from: 1 to: 1312

```

227 ACCTGTCATCCAGCGGCGCCATTAAAGAAAT...TCGGCTACATT 273
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
515 ThArgLYLysSerArGArGLuArGLysArGserGLyAlaAlaYsG1 531
274 GTCGCGCTTCGATCAGCGGCGAGAGTCATCCCGCTTCGACAAACA 323
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
531 uAlaAlaSerSerSerSerGLyThrInProAlaProAlaProAla 548
324 TGCGT.....CACATTCGATTCGTATGAAGCGGTAGTC 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
548 ePrProTrpAspSerLYSLysHISArgSerArGAspArgYsProGLys 564
359 CCGTTGACGATTCAGCCTTACCGCATTCGAGCGGATACAGACAC 408
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
565 HIsAlaSerSerSerAlaArGArGLysArGserArGserArGserA 581
409 CATCCGCGCGGCTATGACGCGGCGAGCGGCTATCCCGCTCC 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
581 rGserThrArGLysArGLysArGLysThrAspArgArGLysGLys 597
459 CAAGGCGCGAGGATATATACGATACGACATTAAGCGGTGCCA 508
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
598 ArGLysSerArGLysArGLysArGLysArGLysArGLysArGL 614
509 ATATCCGCGCT.....CAACGTGACCGACACCGGACGAC 543
    : ||||| :|||:|||||:|||||:|||||:|||||:|||||:
614 rProProProAlaThrSerSerSerSerSerSerArGLuArGLysA 631
544 GGACACAG.....GCTTGTCAGACCGTTTCACAATAC 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
631 rGLyLYSLysHISArGLysGLysSerLYSLysLYSLysArGLys 647
576 CGGTAGTATGCTGACGCAAGAGTAGGAGGATCAAGCCGCGAC 625
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
648 ArgSerArGLyGLyArGLysArGLysGLysArGLysArGLysAla 663
626 GATACAGCCCGAGCTGACAGATCGGCAATCGCGAGAGCTTTTCA 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
664 AlAProAlaPro.....ProProSerGLysThr 673
676 GGCACCTGAGATATCGTCAAAAACATCTCGCGCGCGAGCAAAATT 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
673 rSer.Cys.....GLysArGLysArGLys 680

```

```

726 CGGCGAGCGGATGCCGTGACGAGGTATAAGCAAGCTCAACATTGCTG 775
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
681 ArgArgArg..... 683
776 TTATGACAGCGCTGGGTCTGCTTCCACCGAAACAGAT...GCCGCGC 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
684 .....GLyAlaValProProSerILEGLInAspLeuThrAspH 696
823 ATCAACGATTTCG.....AGATATGGCGCAACCAACGACTATCCGC 866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
696 iAspLeuPheAlaLeuLysArGLyThrIleThrValGLyArgLeuAsp 712
867 ACCACCATCCGCGATTGGGCGAGTCCAAACCCGATCCCGCAGCAAGCA 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
713 SerAsp...ProArgGLyProSerProAlaPro.....AlaSer 724
917 TAGAAGCGGTGACCAATATCTTAC..... 941
724 rSerProLYsArGLyValLeuTYrAspSerGLyGLyLeuSerGLyGL 741
942 .....GGCAGTCATCCCGTCAAAGGAGATGGAGCTGTCCGGG..... 980
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
741 LuArgGLyGLyLysSerSerGLInLYsAspArgArGLysrGLyAlaAla 757
981 ...AAATACGCTTGGCGGCATCAGCGCACATCTGTCAAGCGGTCC 1027
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
758 SerSerSerSerSerSerArGLyGLySerArGLyGLyAlaLeu... 773
1028 AGATGGCGGAGATGTCATTCGCGAAGGGAATCCGCGTACGAGACAAT 1077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
774 .AspLYLysAspArgAspArgAspArgAspArgAspArgAspArg 789
1078 TTTGCCGATGGCGCATACCAAAATACCGCTCCCTTACGATTCCGAAA 1127
789 ..... 789
1128 TATCCGTTCAAACTTGAGAGCGGTACGGCAAAAGCAATCACCCT 1177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
790 ...ArgSerSerLYSLysAlaArgProProLYSGlSerAlaProSer 805
1178 CAACCGTCCGCG.....TCAACGGAAGAATGTGAACCTG 1215
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
805 eArgLYProProLYsProProValSerSerGLysrGLysrSerSer 821
1216 GCAACAAACCCGACCGGACCAAGTCCGTTTACGCTAAAGGTT 1265
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
822 SerSerSerSerSerSerArGLyValLYsLeuInSer..... 835
1266 TCCGAATTTGAAAAGACGTAAATATGATACGAGAAATTAATACGCTG 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
836 .....LYsValAlaValLeuILEArgGLy 845
1316 TACCACAAGTAATCTATAGATGAACCGCTTTATCTCAAGGTTC 1365
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
845 AlSerSerThrProAlaLYsAspAlaSerAlaGLyLeuGLys 861
1366 GTCGATCGGCTCATTTGCTCTATACGCCAGATTCATACGCAAA 1415
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
862 iLeGLy.....ValLYsPheSerAr 868
1416 ATTACCAAGCAAGTAGAATCAGATATATCCACCTAAATAATCTTC 1465
    : ||||| :|||:|||||:|||||:|||||:|||||:|||||:
868 g...AspArgGLysrGLysrProPheLeuLYsProAspGLyArGL 884
1466 CT.....TCAGACCCCTTACCAAAAGACCTAATATGA 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
884 rOThrGLuMetAlaLYsAlaAlaProGLySerThrLYsProLYsLYs 900
1501 TATTGTGAT...AAATTGCTAATGAATGAGCTAAAGT..... 1536
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
901 LYsValLYsAlaLYsAlaGLyAlaLYsLYsThrLYsGLyThrLYsGLy 917

```

```

1537 .....CCATCAGAACTAA..... 1551
1538 |||||:||||:
1539 sThrLysProSerLysThrArgLysValArgSerGlyGlySer 934
1540 .....GGTCAGAAATTGAATGGATGTGTCATGTCTAAACA 1590
1541 1552 .....|||:||||:
1542 934 LysGlySerGlyGlyGln.....ValSerLeuLysLysSer 945
1543 1591 .....CGAAGAGACGCACT 1604
1544 946 LysAlaLysSerCysSerGlnAlaIaGlyThrLysGlyValaGlnGluThr 962
1545 1605 TGGATGGCTAGTAGGAT 1623
1546 962 rSerTrpSerCysGlyGln 968

seq_name: sp_human:Q9H7N4
seq_documentation_block:
ID Q9H7N4 PRELIMINARY; PRT; 1343 AA.
AC Q9H7N4;
DT 01-MAR-2001 (TREMBLrel. 16, created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE FLJ00034 PROTEIN (FRAGMENT).
GN FLJ00034.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC rISSUE-SPIEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024444; BAB15734.1;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
FT NON_TER
FT 1
SQ SEQUENCE 1343 AA; 142483 MW; FA7ABD473C74688C CRC64;

```

```

alignment_scores:
Quality: 145.00 Length: 543
Ratio: 0.604 Gaps: 30
Percent Similarity: 44.199 Percent Identity: 24.125

```

alignment_block:

US-09-303-518d-465 x Q9H7N4 ..

Align seg 1/1 to: Q9H7N4 from: 1 to: 1343

```

227 ACCCTGTCATCCAGCAGGCGCCATTAAAGAAATA...TCGGCTACAT 273
|||:|||||:||||:||||:
546 ThrArgLysLysSerArgArgGlnArgLysArgSerGlyGlnAlaLysGln 562
274 GTCCGCTTTCCGATCAGCGGCGCAGCAAGTCATCCCTTCGACACCA 323
|||:|||||:||||:||||:
562 ValAlaLysSerSerSerGlyThrGlnProAlaProAlaProAla 579
324 TGCCT.....CACATTCGATTCGATGAGCCGCTAGTC 358
|||:|||||:||||:||||:
579 erProTrpAspSerLysLysHisArgSerArgAspArgLysProGlySer 595
359 CCGTTGAGGATTCAGCGCTTACCGCATTCGAGCATGACCAACAC 408
|||:|||||:||||:||||:
596 HisAlaSerSerSerAlaArgArgSer.ArgSerArgSerArgSerA 612
409 CATCCGCGCAGCGCTATGACGCGCCACAGGCGCGGCTATCCGCTCC 458
|||:|||||:||||:||||:
612 rSerThrArgArgArgSerArgSerThrAspArgArgArgGlyGlySer 628

```

```

459 CAAGGCGCGAGGATATATACAGCTACACATTAAGAGCGCTGCCCAA 508
|||:|||||:||||:||||:
629 ArgArgSerArgSerArgGlnLysArgArgArgArgArgSerAla 645
509 ATATCCGCT.....CAACCTGACCGACACACCGACAC 543
|||:|||||:||||:||||:
645 rProProAlaThrSerSerSerSerSerSerArgArgGlnArgHisA 662
544 GCACAAACG.....GCTTGACGCGCTTCCACCAATAC 575
|||:|||||:||||:||||:
662 rGlyLysHisArgAspGlyGlyGlySerLysLysLysLysArgSer 678
576 CGGTAGTATGTCAGCAGAGAGTAGCGGAGGATTCAGACGCGCACCC 625
|||:|||||:||||:||||:
679 ArgSerArgGlyGlnLysArgSer.GlyAspLysSerLysAla.Pro 694
626 GATACAGCCCCGAGCTGCAGACATCGGCGCAATGCCCGCAAGCTTCAAC 675
|||:|||||:||||:||||:
695 AlaProAlaPro.....ProProSerGlySerThr 704
676 GGCACCTGCAGATATGTCAAAACATCATCGGCGGAGGAGAAATGT 725
|||:|||||:||||:||||:
704 rSer.Cys.....GlyAspArgAspSer 711
726 CGGCGCAGCGCATGCCGTCAGAGGTATAAGCGAGCTCAACATTCGTG 775
|||:|||||:||||:||||:
712 ArgArgArg..... 714
776 TTAATGACAGCGTTGGGCTGCTTCACCGAAACACAGAT...GGCGGCG 822
|||:|||||:||||:||||:
715 .....GlyAlaValProProSerIleGlnAspLeuThrAspH 727
823 ATCAACGATTTGGC.....AGMFAATGGCGCAACCAACAGTATGCGCC 866
|||:|||||:||||:||||:
727 IsAspLeuPheAlaIleLysArgThrIleThrValGlyArgLeuAspLys 743
867 AGCAGCATCCCGCATTTGGGCGAGTCCAAAACCCCAATGCGCGCACAGCA 916
|||:|||||:||||:||||:
744 SerAsp...ProArgGlyProSerProAlaPro.....AlaSe 755
917 TAGAAGCCGTCAGCAATATCTTAC..... 941
|||:|||||:||||:||||:
755 rSerProLysArgGlnValLeuLysAspSerGlnGlyLeuSerGlyGlnG 772
942 .....GGCAGTCATCCCGTCGCAAGGAGATTGGAGCTTCGGGG..... 980
772 LysArgGlyGlyLysSerSerGlnLysAspArgArgArgSerGlyAlaIa 788
981 .....AAATACGCGCTTGGGCGCATCAGCGCACATCTGTCAGAGCGTCCG 1027
|||:|||||:||||:||||:
789 SerSerSerSerSerSerArgGlnLysGlySerArgArgLysAlaLeu.. 804
1028 AGATGGCGCAGATCGCATTCGCAAGGCAAAATCCGCGTCAGCGGCACAT 1077
|||:|||||:||||:||||:
805 AspGlyGlyAspArgAspArgAspArgAspArgAspArgAspArgAsp 820
1078 TTTCGCGATCGGCGCATACGCCAAATACCGCTCCCTTACCATCCCGAAA 1127
820 ..... 820
1128 TATCCGTTCAACTTGGAGCAGCGCTTACGCGCAAGAAACATCACTCCT 1177
|||:|||||:||||:||||:
821 ....ArgSerSerLysLysAlaArgProProLysGlnSerAlaProSer 836
1178 CAACCGTCCGCGCG.....TCAACGGAAGAAATGTGAACAG 1215
|||:|||||:||||:||||:
836 erGlyProProProLysProProValSerSerSerGlySerGlySerSer 852
1216 GCAACAAAGCCACCGCAGACCAAGTCCGTTGACGCTGAAGAGGTT 1265
|||:|||||:||||:||||:
853 SerSerSerSerSerSerArgLysValLysLeuGlnSer..... 866

```

```

1266 TCCGAATTTTAAAAAGACGTAAATACGATACGAGATTAATACCGCTG 1315
867 .....LysValAlaValLeuIleAlaGluGly 876
1316 TACCAAGTGAATCTATAGATGAACCGCTCTTAATCTTAAGGTTCT 1365
876 alserSerThrThrProAlaLysAspAlaAlaSerAlaGlyLeuGlySer 892
1366 GTCGATCCGCTCATCTTGTCTTAATACGACGATTCATACGCAAA 1415
893 IleGly.....ValLysPheSerAr 899
1416 ATTACCAAGCAGTACGATCAGATATATCCACCTAAATTAATCTCTC 1465
899 g...AspArgLysSerProPheLysProAspGluArgAla 915
1466 CT.....TCACGACCGCTACCAAAAGACCTAATATATGGA 1500
915 rOThrGluMetAlaLysAlaAlaProGlySerThrLysProLysThr 931
1501 TATTGGAT...AAATTGTAATGAATGACTAAAGT..... 1536
932 LysValLysAlaLysAlaGlyAlaLysLysThrLysGlyThrLysGly 948
1537 .....CCATCAAGACTAAA..... 1551
948 sThrLysProSerLysThrArgLysLysValArgSerGlyGlySerG 965
1552 .....GCTCAAGAAATTGAATGATGATTCATATGCTCAAAACA 1590
965 LysLysSerGlyGlyLysLys.....ValSerLeuLysLysSer 976
1591 .....GGAGAGAGCAACT 1604
977 LysAlaAspSerCysSerGlnAlaAlaGlyThrLysGlyAlaGluGlu 993
1605 TGGATGGCTAGTAGGAT 1623
993 rSerThrSerGlyGluGlu 999

seq_name: sp_human:043418

seq_documentation_block:
ID 043418 PRELIMINARY; PRT: 513 AA.
AC 043418:
DT 01-JUN-1998 (TRENBLREL. 06, Created)
DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)
DE INTERSTITIAL MUCIN (FRAGMENT).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476275; PubMed=9334251;
RA Gum J.R., Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vinall L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
RL terminus, and a novel upstream repetitive region."
RL J. Biol. Chem. 272:26678-26686(1997).
DR EMBL; AF007190; AAC02268.1;
FT NON_TER 1
FT NON_TER 513
SEQUENCE 513 AA; 52573 MW; 33BAC2AFABE1436 CRC64;

```

alignment_scores:

Quality:	144.50	Length:	436
Ratio:	0.695	Gaps:	18
Percent Similarity:	47.706	Percent Identity:	22.018

alignment_block:
us-09-303-518d-465 x 043418 ..
Align seg 1/1 to: 043418 from: 1 to: 513

```

92 CTTTATCCGGCAGGTTCTGACCGCTGACGATTTGAAACCGAGGGA 141
66 LeuValThrThrLeuLeuThrThrThrThrThrThrThrThrThrThr 82
142 TACACCTATTCGACGAGGCGGAACTGCCAGGCGACCGGTCTAT 191
82 uThrThrThrProThrSer.....ProThrSerThrLeuVal 94
192 CGGATTTGGAA.....ACATACAAAGCATCACTAGTGGCAACCTG 232
94 eArpSerThrThrGluThrLysSerThrSerThrThrThrThrThrThr 110
233 TCATCCAGCAGCGCGCCATTAAGAAATATCGGCTACATGTCGCTT 282
111 SerThrAla.....ThrThrLeuProThrThrSerSerSer 123
283 TCCGATCCGCGGACGAGATCCATTCCTCCCTGACACACCATGCTCACA 332
123 uProThrThrGluThrAlaThrMetThrProThrThr..... 136
333 TTCGATTCGATGAACCGGCTAGTCCGTTGACGATTCAGCCTTTACC 382
137 .....LeuThrThrThrThrProAsnThrThrSerThrThr 149
383 GCATCCATTTGGAGCAGATACGACCATTCGCGCAGCGGATGACGG 432
150 ProSerPhe.....ThrSerThrThrThrThrThrThrThrThr 164
433 CCACAGCGCGCGCTATCCCGTCCCAAGCGCGCAGGATATATACAG 482
164 rThrThrAlaThrThrThrThrThrThrThrThrThrThrThrThr 181
483 CTACG.....ACATAAAGCGCTGCCCAAAATATCCGCCCAACCTGA 526
181 eThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 197
527 CCGACACAC.....GCAGCAGCGGACCAACGCGCTTGCACCGT 564
198 ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 214
565 TTCACA..... 571
214 rAlaThrAspLeuThrSerThrPheThrValSerSerSerSerAlaMet 231
572 .....ATACCGTAGTATCTGACGCAAGAGAGGCGACGAT 610
231 eLysSerValThrProSerSerProSerThrThrThrThrThrThrThr 247
611 TCA.....AACGCGCACCGGATACAGCCCGGACCTG..... 643
248 SerLeuValSerMetThrSerAlaThrThrProSerLeuArgProThr 264
644 .....ACAGATGGGCAATGCCCGCAAGCTT..... 670
264 eThrSerThrAspSerThrThrThrThrThrThrThrThrThrThrThr 281
671 .....TCACGCGCTGACGATATGCTCAAAAACATCATGTCG 706
281 eThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 297
707 GCGCGCAGGAGAAATTTGCGCGCAGCGGATGCGGCTGACGATTAAC 756
298 HisThrGluThrThrThrThrThrThrThrThrThrThrThrThrThr 314
757 GAAGGCTCAACATTTGCTGATATGACAGCGCTGGGCTGCTT..... 799
314 rThrAlaGluSerAlaLeuAlaProThrThrThrThrThrThrThrThr 331

```

```

800 .....CCACGAAACAGATGGCGGCATCAAGATTGG 835
      |||||
331 erProthMetgluProProSerThr.....ThrValAla 342
836 CAGATATGGCGCACTCAAGACTATGCGCGACGACCATCGCGCATGG 885
      |||||
343 ThrThrGlyThrGlyInThrThrPheProSerSerThrAlaThrPhe 359
886 GCGTCGCAAAACCCCAATGCGCGACGACGATGAAAGCGTCGCAATAT 935
      |||||
359 uGluThrThrThrLeuThrProThrThrAspPheSerThrGluSerLeu 376
936 CTTTACGCGATCATCC.....CCGTCA 958
      |||||
376 hrThrAlaMetThrSerThrProProIleThrSerSerIleThrProThr 392
959 AAGGATGTAGAGCTGTGGGAAATATGCGCTTGGCGCATCAAGCA 1008
      |||||
393 AspThrMetThrSerMetAlaThrThrThrSerThrProThrAlaThrAs 409
1009 CATCTGTCAAGGCGTCGACATGGCGAGATCGCATGGCGCAAGGAA 1058
      |||||
409 nThrLeuSerPro.....LeuThrSerSerIleLeuSerSerT 422
1059 ATCCGCGCTGAGGACATTTTCCGATGCGCATCGCAATATCCCT 1108
      |||||
422 hrProValProSerThrGluValThrThrSerThrThrAsnThrAsn 438
1109 CCCCTT.....ACCATTCGCGAATATCCGTTCAACTTG 1143
      |||||
439 ProValSerThrLeuValThrThrLeuPro...IleThrIle...ThrAl 453
1144 GAGCAGCGTTACGCGCAAGAAACATCACTCTCTCAACCGTGGCGCGCT 1193
      |||||
453 gSerThrLeuThrSerGluThrAlaThrProSerSerProThrSerThrV 470
1194 AAACGGAA 1201
      |||||
470 alThrGlu 472

```

seq_name: sp_human:014881

```

seq_documentation_block:
ID 014881 PRELIMINARY; PRT; 622 AA.
AC 014881;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MUCIN (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEOBRONCHIAL MUCOSA;
RA Dessevyn J.L., Guyonnet-Duperrat V., Porchet N., Aubert J.P., Laine A.;
RT "Human mucin gene MUC5B: the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super repeat.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X74955; CAA52910.1; -.
FT NON_TER 1
FT NON_TER 1
FT SEQUENCE 622 AA; 61786 MW; 4FC85A52F50D57E5 CRC64;

```

alignment_scores:

Quality:	144.00	Length:	418
Ratio:	0.738	Gaps:	18
Percent Similarity:	46.651	Percent Identity:	23.923

alignment_block:

```

US-09-303-518D-465 x 014881 ..
Align seg 1/1 to: 014881 from: 1 to: 622
68 ACAGCTCAGATTGGCAACAGATTCTTTATCCGGCAGGTTTCGACCG 117
      |||||
187 ThrProGluThrThrHisThr.....SerThrVal 196
118 CAGATTTCGACACCCGCGGAAATACACCTATTCGCGACGAGGGGA 167
      |||||
196 ThrThrThrThrAlaThr.....ThrThrAlaThrSerThr 211
168 ACTGCGCGAGCGCGATATCGGATGGGAAACATCAAGCCATC 217
      |||||
211 erThrProSerSerThrProGlyThrThrThrIleuThrGluLeuThr 227
218 AGTTGGGACCTGTCTCATCCAGCGCGCA..... 250
      |||||
228 ThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSe 244
251 .....TTAAGGAAATATCGGCTACATTTGCG 278
      |||||
244 rThrProGlyThrThrThrIleuThrGluLeuThrThrThrAlaThr 261
279 CTTTCCGATCAGGCG.....ACGAGTCATTCCTCCGTCGACA 319
      |||||
261 hrThrAlaSerThrGlySerThrAlaThrLeuSerSerThrProGlyThr 277
320 ACCATGCTCATTCGATTCGTATGTAAGCGGTAGTCCGTTGACGGA 369
      |||||
278 ThrTrp.....IleuThrGluProSerThrThrAlaThrVal 290
370 TTCACCTTTACCGCATTCGATTTGGAGCGGATACGACCATCCCGCGA 419
      |||||
290 ThrValProThrGlySerThrAlaThrAlaSerSerThr...GlnAla 306
420 CCGCATGACGGCGCACAGGCGCGCTATCCGCTCCCA..... 460
      |||||
306 hrAlaGlyThrProHisValSerThrThrAlaThrThrProThrValThr 322
461 .....AAGCGCGAGGATATATACAGCTACGACATAAAGCGTGGC 504
      |||||
323 SerSerThrAlaThrProSerSerSerProGlyThrAlaThrAlaLeu 339
505 CAAATA...TCCGCTCAACCTGACCGCA.....ACCGAG 539
      |||||
339 alAlaLeuArgSerThrAlaThrThrProThrAlaThrSerPheThrAla 356
540 CACCGGACACGCGTTGTCGACCGTT.....TCCACA 571
      |||||
356 leProSerSerSerLeuGlyThrThrThrArgLeuSerGlnThrThr 372
572 ATACCGGTATGTGTCGACGCAAGAGTAGGCGAGGATCAAGCGCGC 621
      |||||
373 ThrProThr.....AlaThrMetSerThrAlaThr 382
622 ACCGATACAGCCCGAGCTGCGACAGATGGCGCAATGCCCGAAGCTT 671
      |||||
382 rProSerSerThrProGluThrVal.....HisThr 393
672 CAACGGCACTGCAATATCGTCAAAACATCATCGCGCGCGGAGGAA 721
      |||||
393 erThrValLeuThrAlaThrAlaThrThrThrGlyAlaThrGlySerVal 409
722 TTGTCGCGCGAGCGCATGCCGTCGAGGATTAAGCAAGCTCAACATT 771
      |||||
410 AlaThrProSerSerThrProGlyThrAlaHisThrThrThrValProth 426
772 GCTGTATGCAAGCGCTTGGCTGCTTTCACCGGAAACAGATGCGCG 821
      |||||
426 rThrThrThrThrGlyPheThrAlaThrProSerSerSerProGlyThrV 443
822 CATCAACGATTGGCAGATATGGCGCACTCAAGACTATGCCG..... 865

```



```

1692 yalAla1a1aNet.....AlaSerThrLeuSerSerProleu 1703
626 GATACAGCCCCGAGCTGGACAGATCGGCGAATG..... 658
1704 LysGlnMetSerCylHisAlaGluValAlaLeuValAsnGlySerValSe 1720
659 .CCGCCGAGACTT.....TCACGGCAGCTGCAGATATCTCAAAACAT 701
1720 rProLeuLysTyrProSerSerSerAlaLeuIleAsnGlyCysLysAlaT 1737
702 CATCGCGCGGCGAGGAGAAATTTGTGGC.GCAGGCGATCCGTCGACGCT 750
1737 hAlaThrLeuIleAsnPlsSerThrAlaThrAsnAlaValSerSer 1753
751 ATACGCGAGGCTCAACATTCCTGTATGCACGCGCTGGGTCTGCTTC 800
1754 ValValSerAlaAlaProAspThrValGluLysAla.....LeuSe 1767
801 CACGCAAAACAAGATGGCGCCCATCAACGATTTGGCAGATGGCGCAAC 850
1767 rThrThrThrAlaMet.....ProheserProL 1777
851 TCAAGACTATCGCGACGACGACCATCGCGATTTGGCAGTCCAAACCCC 900
1777 euArgSerTyrValSerAlaAla.....Pro 1786
901 AATGCCGCACAGGCATAGAACCC.....GTCAACAATATCTTTACGCG 944
1787 SerAlaPheGlnSerLeuArgAlaProSerAlaSerAlaLeuTyrAsnSe 1803
945 AGTCATCCCGCTCAAGGAGAT.....GAGCTGTTCGGGAAAA.... 984
1803 rLeuGlyProSerValGlyValThrThrSerSerValThrSerSerLeu 1820
985 .....TACGGCTTGGCGCGCATCAAGCGGCGACATCCGTCAG 1020
1820 LeuThrValProValTyrSerValGlyAsnValLeuAlaGluProAlaLeu 1836
1021 CGGTCGAGATGGCGCGAGATCGCATTTCCGAAAGGAATCCGCGCGAC 1070
1837 LysLysLeuProAspSerAsnSerLeuThrLysSerAlaAlaLeuLe 1853
1071 GCACAAATTTTGGCGATCGGCATACGCCAATACCGTCCCTTACCAT 1120
1853 uSerProIleLysThrLeuThrGluThrArgProGlnProIlePheA 1870
1121 CCGCAAT.....ATCCGTTCAACTTTGGAGCAGCGTTACGGCAAA 1161
1870 snArgThrSerSerProValLysSerSerLeu..... 1880
1162 GAAACATCTACTCTCAACCGCGCGCATCAAGCGAAAGAAATGTCAA 1211
1881 ..PheLeuAlaSerSerAlaLeuLysProSer..... 1890
1212 ACTGGCAACAAAGCCGCCGGAAGACCAAAAGTCCGTTTGAAGTAAAG 1261
1891 .....ValProSerSerLeuSerS 1897
1262 GATTTCGAAATTTGAAAAAGAGTA.....AAATACGATACGAGA 1302
1897 eArgGlnGluIleLeuLysAspValAlaGluMetLysGluAspLeuMet 1913
1303 ATTAATACGCGCTGTACACAGATGCTATAGAACCCGTCCTTTAA 1352
1914 ArgMetThrAlaIleLeuGlnThrAspValProGluGluLysProPheI 1930
1353 TCTTAAGGTCTGTGGATCGGCTCATCTTGTGCTATACTGCCAGAA 1402
1930 ..... 1930
1403 TTCAATACGCAAAATTACACAGGAGTATGATGATATATCCACCT 1452
1931 .....ThrAspLeuProAlaGluGlyArgIleAspAspGluPro 1944

```

```

1452 ..... 1452
1945 PheLysIleValGluLysValLysGluAspLeuValLysValSerGluI 1961
1453 .....AAAAATACCTCTCTCAGACCGCTACCAAAAGACCTAATA 1495
1961 eLeuLysLysAspValAlcLysValGluSer.....LysGlyPro.... 1973
1496 ATGATATTGTAATAATTGTAATGATGACTTAAGGTCATCGATCAAGA 1545
1974 .....ProlsSerProLysSer 1979
1546 ACTAAGCTAA.....GAATTTGAATGGGATGTTCAATTTCTAAGAC 1589
1980 AspLysGlyHisSerProGluAspAspTrpThrGluPheSerSerGlu 1996
1590 AGGAGAGAGCACTTGATGGGCTAGTAGGATGGTGAAGCATTTAATA 1639
1996 uIleArgGluAlaArgGlnAlaAlaAlaSerHisAlaProSerLeuPro 2013
1640 TATCAATTGATGGAAAG 1656
2013 LuArgValHisGlyLys 2018

seq_name: sp_bacteria:Q53478

seq_documentation_block:
ID Q53478 PRELIMINARY; PRT; 456 AA.
AC Q53478;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DR 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ORF2 PROTEIN.
CM ORF2.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95020543; PubMed=7934834;
RA Facius D., Meyer T.F.;
RT "A novel determinant (coma) essential for natural transformation
RT competence in Neisseria gonorrhoeae and the effect of a coma defect on
RT pilin variation."
RL Mol. Microbiol. 10:699-712(1993).
DR EMBL: S75490; AAB32262.1;
SO SEQUENCE 456 AA; 54380 MW; 30ACT7B5CDBF912D5 CRC64;

alignment_scores:
Quality: 141.00 Length: 456
Ratio: 0.691 Gaps: 28
Percent Similarity: 44.737 Percent Identity: 24.342

alignment_block:
US-09-303-518D-465 x Q53478 ..
Align seg 1/1 to: Q53478 from: 1 to: 456

52 CTGCCGATGCATGCACAC.....GCCTCAGATTGGCAGAA 86
1 MetProGlnHisGlnHisLeuAlaValGlnValAlaSerIleLeuAlaAs 17
87 CGATTCTTTATCCGGCGAGTTCTGCACCCGTCAGCATTTGC..... 127
17 pGlnAlaPheLeu.....AlaLeu.ArgIleGlnIleSerArgTyrGln 31
128 .....AACCGAGCGGAATACCACTATTTCGGCAGC 159
32 GlnCysArgThrAlaArgHisAsnThrGlnAsnThrThrPhe..... 45
160 AGGGGGGAACTTGGCGAGCGGCGGTCATATGCGATTGGGAACATATCA 209

```

```

46  ....ProIleIleIleAspIlePheIleAlaPhe. 54
210 AAGCCATGAGTTGGGCAACCTGTTTCATGACACAGCGGCCCATTTAAAG... 256
55  ....ArgAspProGlnIlySerIeu 60
257 GAATATCGGCTACATGTTGCGCTTTTCCGATCCAGGACAGAGTCCAT 306
61 GlnIleHisAla...Va 65
307 TCCCCCTTCGACA...ACCATGCTCCACATTCGATTCG... 343
65  IProIeuProIthrIeuPheArgIthrIeuProIArg...ProIeuIleIySp 81
344  ....ATGAGCGGGTAGAGCCGCTTGAGCGAGTTCAGACCTT 379
81 heArgIeuSerGlyIleAsnProIAlaValArgIyIleAlaIAspCysIeu 97
380 ACCCGATTCATTTGGGAGCGGATAGCAACACCATCCGCGCGAGCTTAGC 429
98 LysAlaIleAlaVal...ValValAlaIleAlaIleMetArg 108
430 GGGCCACAGCGCGCGGCTATCCCGCTCCCAAAAGCGCGGAGGATATATA 479
108 gclAsnIleIleuValGln...ThrAlaAspThrArgArg...Ile 121
480 CAGCTAGACACATTAAGAAGCGT...TGCCCAAAATATTCGCGCTCAACCTGA 526
122 GlnIthrArgHisAspThrArgIeuCysArgIyIThrArgSerArgIleG 138
527 CCGACACCGGACAGCACCGGACGCTTGTTCGACCGCTTCCACAATAC 576
138 uIySerGln...MetAlaIleValCysArgPro... 146
577 GGTATGATGTCAGCCAGAGGATAGCGACGATTCCAAAAGCGCCACCG 626
147  ....HisIyHisIleArgGlnAlaIleuProArgIlePro 157
627 ATACAGCCCGGACGTGACAGATCGGGACATGCCCGCGACCTTT... 671
158 AsnArgAsnArgSerPheValArgPheArgIyCysPheArgIThrValGlyTy 174
672  ....CAAGGGCAC...TGCAG 684
174 TlySerProCysGlnGlnGlnHisArgProArgIThrGlnIthrGlnAlaIle 191
685 GATATCGTCAAAAATCATCATCGCGCGGACAGAAATTTGTGCGCGCAG 734
191 rglGlnGlnGlnGlnGlnIleSerGlyIThrHisArgGlnHisGlnIleArgGln 207
735 CGATGCGCGTGCAGGGTATAGGGAAGGCTCAAAACATTCG...TGTTA 778
208 ArgCysGlyGlyAsnGlyIlySerIeuArg...ArgHisIleGlyIlySerProH 223
779 TGCAGCGGCTTGGGTGCTGCTTCCACCGAACAACAGATGCGCGGCATCAAC 828
223 scGlnIySerProGlnSerValArgPheArgGlnIlySerCysProArgIleuIleu 240
829 GATTTGGCAGATGGCGCACTCAAAAGCATATGCCCGCAGACCGCATCCG 878
240 rglArgIyAspArgHisAspArgIThrGlnIleArgGlnIleArgIlyGlnHisPro 256
879 CGATTGGGAGATGCCAAAAGCCCAATGCCGACAGAG... 914
257 Arg...IlySerArgIyIThrArgHisArgIleAspGlnArgAl 269
915  ....C 915
269 AspArgIleArgGlnArgGlyIlySerGlnIleThrGlnGlnIleAsnGlnIleArgGln 286
916 ATTAGAAGCGGTACAGATAATCTTTACGGCGAGCGTATCCCGGTCAAGAGAT 965

```

```

286  isAtgserAtgProLeuProAlaHisIsgLYgLYLeuProPheProLeuVal 302
      ::::: ::::: ||||| ::::: ||| :::::
966  TGGAGCTTGTGGGGAAATATACGGCTTGGCGGCATACGGCAATCTCTG 1015
      ::: ::::: ||||| ::::: ||| :::::
303  GlnThrProCysThrArgProTyrGlnIleYargHisGln...ThrGluAr 318
      ::::: ::::: ||||| ::::: ||| :::::
1016  TCACACGGTCCGACAGATGGGGCAGATCCG...ATTGGCCAAAGGAAATCC 1062
      ||| ::::: ||||| ::::: ||| :::::
318  gGlnProSerThrAspTrnGlnAsnArgGlnArgValGlnGlnGlnHisA 335
      ::::: ||||| ::::: ||| :::::
1063  GCCGTCACAGCAAAATTTTCCGATGCGGCATACGCCAA...ATACCCGCTC 1109
      ::::: ||||| ::::: ||| :::::
335  tGArgGlnArgLeuProProAlaGArgProArgArgGlnSerSerProAla 351
      ::::: ||||| ::::: ||| :::::
1110  CCCTTACCATTCGCCAAATATCCGTTCAAACTTGGACGACGGTTACGGCA 1159
      ||||| ::::: ||||| :::::
352  ProArgPro.....ArgGln 356
      ::::: ||||| ::::: ||||| :::::
1160  AAGAAACATTCACCTCTCTCAACCGTGGCGCGCTCAACGGAAGAAATGTG 1209
      ::||| ::::: ||||| ::::: ||| :::::
356  uArgArgGlnHis.....GlnAsnThrAlaLeuArgArgHisGLYL 370
      ::||| ::::: ||||| ::::: ||| :::::
1210  AAACGTGGCAAAACA 1223
      ::||| ::::: ||||| ::::: ||| :::::
370  ystHrGlyLysGln 374
      ::||| ::::: ||||| ::::: ||| :::::

```

```

seq_name: sp_fungi:P78953
seq_documentation_block:
ID      ID      PRELIMINARY:   PRT;      920 AA.
AC      P78953;
DT      01-MAY-1997 (TREMBLrel. 03. Created)
DT      01-MAY-1997 (TREMBLrel. 03. Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17. Last annotation update)
DE      DMF1 GENE.
GN      DMF1.
OS      Schizosaccharomyces pombe (fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC      Schizosaccharomycetales; Schizosaccharomycetaceae;
OC      Schizosaccharomyces.
OX      NCBI_TaxID=4896;
RN      RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=972H-;
RX      MEDLINE=97102672; PubMed=8946912;
RA      Soizmann M., Fankhauser C., Brobeck C., Simanis V.;
RT      "The dmf1/mdl gene is essential for correct positioning of the
RT      division septum in fission yeast.";
RL      Genes Dev. 10:2707-2720(1996).
RN      RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=972H-;
RC      Seeger K., Harris D., McDougall R.C., Rajandream M.A., Barrell B.G.;
RL      Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Y07599; CAA68873.1; -.
DR      EMBL; AL132870; CAB60689.1; -.
DR      InterPro: IPR001849; PH.
DR      Pfam: PF00169; PH; 1.
DR      SMART: SM00233; PH; 1.
DR      PROSITE: PSS0003; PH; DOMAIN: 1
SO      SEQUENCE      920 AA; 102366 MW; CA95FB48A20549B5 CRC64;

```

[illegible]

```
alignment_block:
```

Align seg 1/1 to: P78953 from: 1 to: 920


```

101 GGCAGGTTCTCAGCCGTCAGATTGCAACCCGAGGGAATATCCACCTA 150
|||||: : : : : ||| ||| |||: |||
123 G|A|A|G|I|E|P|R|O|S|E|L|E|U|T|H|Y|S|E|A|L|G|I|N|H|E|T|A|S|P|R|G|I| 138
|||||: : : : : ||| ||| |||: |||
151 TTCGCGACGAGGGGAACTTCCGACGCGACGGGTTCATATCGATTGG 200
|||||: : : : : ||| ||| |||: |||
139 .SerSerSerThrAlaSerLeuLysG|Y|A|L|A|S|P|L|E|V|A|L|A|S|P|R|G| 155
|||||: : : : : ||| ||| |||: |||
201 AAACATACAAA.....GCCATCAGTTGGGCAACCTGT 232
|||||: : : : : ||| ||| |||: |||
155 L|U|T|H|P|H|E|A|S|P|R|O|A|S|P|I|N|S|L|E|U|A|L|G|I|L|E|U|S|E|R|P|H|E|A|S|P|R|G|S|R 171
|||||: : : : : ||| ||| |||: |||
233 TCATCCAGACGCGG.....CCATTAAAGAAATATCGGCTACACTT 273
|||||: : : : : ||| ||| |||: |||
172 LysSerSerArgLysArgAlaValAluValAlaGluPheSerAlaGlyLys 188
|||||: : : : : ||| ||| |||: |||
274 GTCCGCTTTCCGATCAGCGGCGCAAGTCCATTTCCCTTCGACA... 319
|||||: : : : : ||| ||| |||: |||
188 sThiMetSerProLeuGluTyrThrValGlnHisProTyrGlnSerHisA 205
|||||: : : : : ||| ||| |||: |||
320 .....ACC 322
|||||: : : : : ||| ||| |||: |||
205 sngL|U|L|E|U|S|E|R|T|H|A|S|P|R|O|A|L|A|A|A|A|A|G|L|Y|S|E|R|V|A|L|P|R|O|A|S|N 221
|||||: : : : : ||| ||| |||: |||
323 ATGCTTCATTCGATTCGATTCGATGAAGCGGTAGTCCGTTGACGATTC 372
|||||: : : : : ||| ||| |||: |||
222 L|E|U|A|A|G|I|L|E|P|R|O|S|E|R|A|S|P|V|A|L|Y|S|P|R|O|A|L|H|I|S|L|E|U|S|E 238
|||||: : : : : ||| ||| |||: |||
373 AGCCTTTCACGATTCATTTGGAGGAGTACGAACCATCCCGCGGACGG 422
|||||: : : : : ||| ||| |||: |||
238 rAla.....SerSerThrValG|Y|P|R|O|A|G|I 247
|||||: : : : : ||| ||| |||: |||
423 CTATAGCGGCGCCACAGCGCGGCTATCCGCTCCCAAGGCGCGAGG 472
|||||: : : : : ||| ||| |||: |||
247 L|E|U|P|R|O|S|E|R|.....L|E|U|P|R|O|L|Y|A|S|P|... 254
|||||: : : : : ||| ||| |||: |||
473 ATATATACGCTACGACATAAAGCGCTGCCCAAAATATCCGCTCAAC 522
|||||: : : : : ||| ||| |||: |||
255 .....Th|H|I|G|L|U|A|S|P|A|S|P|R|O|A|L|E|U|L|A|A|V|A|L|G|I|N|... 266
|||||: : : : : ||| ||| |||: |||
523 CTGACGACACGCGACA.....CCGACACAGCGT 554
|||||: : : : : ||| ||| |||: |||
267 .....Th|H|I|A|S|E|R|L|E|U|A|S|P|E|L|A|S|P|Y|L|Y|S|P|R|O|L|E|U|G|I|N|P|R|O|L 281
|||||: : : : : ||| ||| |||: |||
555 TGTGACCGTTCCACAATACCGGTACTAGCTGACGCAAGAGTAGGCG 604
|||||: : : : : ||| ||| |||: |||
281 e|U|A|I|A|P|R|O|L|E|G|I|N|L|A|I|A|P|R|O|A|L|..... 289
|||||: : : : : ||| ||| |||: |||
605 ACGGATTCAACGCGCCACCGCATACAGCCGAGCTGACAGATGGGCG 654
|||||: : : : : ||| ||| |||: |||
290 G|L|U|A|S|P|H|S|E|R|G|I|N|P|R|O|H|E|S|E|R|S|E|R|V|A|L|P|R|O|L|U|A|L|H|I|S|T|H|R 303
|||||: : : : : ||| ||| |||: |||
655 AATGCGCGCAACGCTTTCACGCGACTGCAGATATGTCAMAAACATCAT 704
|||||: : : : : ||| ||| |||: |||
304 .....L|E|U|A|S|P|S|E|R|A|S|P|I|L|E|S|E|R|T|H|G 312
|||||: : : : : ||| ||| |||: |||
705 CGCGCGCGCAGAGAAATTGTCGCGCGCATGGCGGATGGCGGAGGATATA 754
|||||: : : : : ||| ||| |||: |||
312 L|U|S|E|R|L|E|U|A|T|G|L|Y|S|V|A|L|L|E|U|A|L|Y|S|E|W|C|I|U|A|L|Y|S|A|G|..I|L|E|S 328
|||||: : : : : ||| ||| |||: |||
755 GCGAAGGCTCAACATTTGCTGTTATGACAGCGCTGGCTGCTTCACCC 804
|||||: : : : : ||| ||| |||: |||
328 e|T|S|E|R|S|E|R|S|E|R|Y|A|L|A|.....S|E|R|T|H|R 336
|||||: : : : : ||| ||| |||: |||
805 GAAACAGATGGCGCGCATCAACGATTTGGCAGATATGGCCCACTCA 854
|||||: : : : : ||| ||| |||: |||
337 L|E|U|A|G|L|Y|S|V|A|L|Y|A|S|P|H|S|E|R|G|I|L|E|U|S|E|R|L|E|U|P|R|O|H|A|S|N|G|L|Y| 353
|||||: : : : : ||| ||| |||: |||
855 AGACTAT.....G 862
|||||: : : : : ||| ||| |||: |||
353 s|A|S|P|Y|A|S|P|G|L|E|U|T|Y|R|L|E|U|G|I|N|S|E|R|A|R|G|A|S|N|S|E|R|G|I|N|P|R|O|G|I|N|T 370
|||||: : : : : ||| ||| |||: |||
863 CCGCAGACCATCCGCGATTGGGAGTCCAAACCCCAATGCCGACGAA 912
|||||: : : : : ||| ||| |||: |||

```

```

seq_name: sp_bacteria:069385
seq_documentation_block:
ID 069385 PRELIMINARY: PRT: 1390 AA.
AC 069385;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DE 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE GLUCOSYLTRANSFERASE-ST.
GN GTF.

```



```

1051 rsnllleuglyatrglylaglytyr.....vall 1062
1289 AATGATATGAGAAATTAATACCGGTGACCAAGTGAATCTTACAT 1338
1062 eullyspdlinalathrasnthr.....TyrPheSerLeuValSer 1075
1339 GAACCCGCTTTAAATCCATAAGGTTCTGCGATCGGCTCAT.....TC 1382
1076 AspAsnThrPheLeuProLysSerLeuValAsnProAsnHisGlyThrSe 1092
1383 TTGGCTATATGCGAGAAATTCATACGCAAAATTACCAGGCAAGTA 1432
1092 rsersevalthrlyleuvalpheaspolylys.....glyt 1105
1433 GAATGATATATCCACCTAAATTAATCTCTCTTACGACCGGTACCA 1482
1105 yvaltyrtyrserthrserglyasnnglnalalyasnalaphelleser 1121
1483 AAGAGCACTAATATGATATTGATTAATTTGTAATGAATGACTAA 1532
1122 leuglyasnasntrpttyrtyrphespasnnglytyrmetvalthr.. 1137
1533 AGTTCATATCAAGAACTAAAGTCAAGATTTGAATG.....GATG 1573
1138 glyalaglnserileasnnglyalasntrtyrtyrphleuSerasnngly 1154
1574 TTCAATTGTCTAAA.....ACAGAGAGAGCAACTTGGATGG 1611
1154 leglleuargasnalailettyrasnnglyasnlysnlyvalleuSerlyr 1170
1612 GCTAGTAGGATGGTAAAG 1629
1171 tyrglyasnaspolyarg 1176
seq_name: sp_bacteria:069391

```

```

seq_documentation_block:
ID 069391 PRELIMINARY; PRT: 1455 AA.
AC 069391:
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
GN GUCOSYLTRANSFERASE-SI.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4251;
RC MEDLINE=98331643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RT FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL: D88661; BAA2614.1.;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 9.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

```

alignment_scores:

Quality:	140.00	Length:	606
Ratio:	0.507	Gaps:	32
Percent Similarity:	45.545	Percent Identity:	22.112

alignment_block:

US-09-303-518d-465 x 069391 ..

```

Align seg 1/1 to: 069391 from: 1 to: 1455
64 GCACAGCCCTCAGATTGGCAAAAGATTCTTTATCCGCGAGGTTCGCA 113
643 SerTyrAlaLeuLeuLeuThrAsnLysSerSerValProAlaValYrTy 659
114 CCGTCACCATTTGCAACCCGAGGGAATAC..... 144
659 rglYAspMetPheThrAspAspGlyGlnTyMetAlaHisGlyThrIleA 676
145 .....CACSRTATTC 153
676 snTyrgluAlaIleGlnThrLeuLeuLysAlaArgIleLysTyValSer 692
154 GGCAGCAGGGGGGAACCTTCCGAGCGCGGTCA..... 189
693 GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGluIleIleTh 709
190 .....ATCGATTGGGAACATACAAACCCATCAGTTGGCAACC 229
709 rSerValArgTyrglyLysGlyValAlaLeuLysAlaThrAspThrGlyAspA 726
230 TGTTCAATCCAGCAG.....GCGGCCATTAAGAAATATCGGCTAC 270
726 rgrThrThrArgThrSerGlyValAlaValIleGluGlnAsnProSer 742
271 ATT...GTCCGCTTTCCGAT.....CA 290
743 LeuArgLeuLysAlaSerAspArgValValAlaAsnMetGlyAlaAlaHis 759
291 CGGCAAGAGATCCATTCCTCCCTTC.....GACAAC.....C 322
759 slyAsnGlnAlaIleArgProLeuLeuLeuThrThrAspAsnGlyIleL 776
323 ATGCTTCACATTCGATTCGATGAGCCGATAGTCCGTTGACGGATTC 372
776 ysaLatyrHisSerAspGlnGluAlaIleGly..... 786
373 AGCCTTTACCGCATTCATGGGAGCATACGAA.....CACCATCCGC 416
787 ...LeuValAlaArgTyThrAsnAspArgGlyGlnLeuIlePheThrAlaAl 802
417 CGAC.....GGCTATGACGGCCACAGCGCGCGCTATCC..... 453
802 asprileLysGlyTyralaAsnProGlnValSerGlyTyrglyLeuGlyValT 819
454 ..GCTCCAAAGGCGCGAGGATATATACAGCTACGACATAAAGCGCTT 501
819 rValAlaProValGlyAla.....AlaAla 826
502 GCCCAAAATATCGCGCTCAACCTGACCGCAACCCGACGCGGCAACG 551
827 AspGlnAspValArgValAlaAlaSerThrAlaProSerThrAspGlyL 843
552 GCTTGTCAGCCGTTTCACAAATACCGTAGTATGTCGACGCAAGAGTAG 601
843 sSerVal.....HisGlnAsnAlaAlaLeuAspSerArgValMetP 857
602 GCGAGGATTCAAACGCCGACCCGATACAGCCCGGAGCTGACGATCG 651
857 hegluclyPheSerAsnPheGlnAlaPheAlaThrLysGluGlnLurYr 873
652 GGCATGCGCGCGAGCTTCAC.....GGCAC 680
874 ThrAsnValAlaIleAlaLysHisValAspLysPheAlaGluTyrGlyVa 890
681 TGCAGATATCGTCAAAACATCATCGCGCGCAGAGAAATTTGCGCG 730
890 lThrAsp.....PheGluMetAlaProGlnIleTyralaIle 902
731 CAGGCGATGCCGTG.....CAGGCTATACCGAAGCTCAACATTT 771

```

```

902 eThraspGlySerPheLeuAspSerValIleGlnAsnGlyTyrAlaPhe 918
772 GCCTGATATGACGGCTTGGTGTCTGCTTCCACGAAACAGATGGCGCG 821
919 ThrAspArgTyrAspLeuGly...IleSerLysProAsnLysTyrGlyTh 934
822 CATCAACGATTTTGGCAGAT...ATGGCGCAACTCAAAAGCTATGCCGCG 868
934 rAlaAspAspLeuValIleLysAlaIleLysAlaLeuHisSerLysGlyIle 951
869 CACCCATCCCGGATTTGGCGAGTCCAAACCCCAATGCCGCAAGGCGATA 918
951 ysaValMetAlaAspTyrValProAspGlnMetLysAlaLeuProGluLys 967
919 GAAGCCGTCACAAATATCTTTACGGCAGTCATCCCGTCAAAAGGATG 968
968 GluValValThr..... 971
969 AGCTGTTGCG...GGAATAATACGGCTTGGCGGCGCATCAGCGCATCTCG 1015
972 .AlaThrArgValAspLysTyrGlyThr.....Proy 982
1016 TCAAGCGTCGACAGATG...GGCGAGATCCGATTCGCGAAGGAAATCC 1062
982 aLaAlaGlySerGlnIleLysAsnThrLeuTyrValValAspGlyLysSer 998
1063 GCCGTCAGCGCAAT.....TTGCCGATCGCGC 1091
999 SerGlyLysAspGlnGlnAlaLysTyrGlyAlaPheLeuGluGlu 1015
1092 ATAGCCCAATACCGCTCCCTTACCATTCGCGAATATCCGTTCAAACT 1141
1015 uGlnAlaLysTyrProGluLeuPheAlaArgLysGlnIle..... 1028
1142 TGGAGCAGCGTTACGGCAAAAGAAACATCACTCTCAACCGTCGCCCGC 1191
1029 .....SerThrGlyValProMet 1034
1192 TCAAACGGAAGAAATGTGAA...CTGGCAAAACAACGCCACCGGAGAC 1238
1035 AspProSerValLysIleLysGlnThrProSerAlaLysTyrPheAsnGlyTh 1091
1239 CAAGTCCCGCTTTGACGGTAAAGGCTTCCGAATTTGAAAGAGCTAA 1288
1051 rAsnIleLeuGlnArgGlyAlaGlyTyr.....ValL 1062
1289 AATAGCATACAGAAATATATACCGCTGACCAACAAGTGAATCTATGAT 1338
1062 euLysAspGlnAlaThrAsnThr.....TyrPheSerLeuValSer 1075
1339 GAACCGCTTCTTAATCTTAAGGTTCTGTCGATCGGCTCAT...TC 1382
1076 AspAsnThrPheLeuProLysSerLeuValAsnProAsnHisGlyTh 1092
1383 TTGCTCTTAACTGCCAGATTCATACGCAAAATTCACCAAGGAGCTA 1422
1092 rSerSerValThrGlyLeuValPheAspGlyLys.....GlyT 1105
1433 GAATGCAATATATCCACCTAAATAATCTCTCCCTTACGACCGCTTACA 1482
1105 yValTyrTyrSerThrSerGlyAsnGlnAlaLysAsnAlaPheIleSer 1121
1483 AAAGACCTTAATATGATATTTGGATAAATTTGGTAATGATGACTAA 1532
1122 LeuGlyAsnAsnTyrTyrThrPheAspAsnGlnGlyTyrMetValThr 1137
1533 AGCTCATCAACAAGCTAAAGTCAAGAAATTTGAATG.....GATG 1573
1138 .GlyAlaGlnSerIleAsnGlyAlaAsnTyrTyrPheLeuSerAsnGlyI 1154
1574 TTCATATTTCTTAAA.....ACAGAGAGAGACCAACTTGATGATG 1611
1154 IeGlnLeuArgAsnAlaIleTyrAspAsnGlyAsnLysValLeuSerTyr 1170

```

```

1612 CCTAGTAGGATGCTAAG 1629
1171 TyrGlyAsnAspGlyArg 1176
seq_name: sp_bacteria:069397

```

```

seq_documentation_block:

```

```

ID 069397 PRELIMINARY; PRT: 1455 AA.
AC 069397;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC SRAIN-MT4467;
RX MEDLINE=98231643; Pubmed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans."
RL EMS: Microbiol. Lett. 161:331-336(1998).
DR InterPro: IPR002479; CW_binding.
DR Pfam: PF01473; CW_binding_1; 9.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW transferase.
SQ SEQUENCE 1455 AA: 162914 MW: A1263427BF24E8E1 CRC64;

```

```

alignment_scores:
  Quality: 140.00 Length: 606
  Ratio: 0.507 Gaps: 32
  Percent Similarity: 45.545 Percent Identity: 22.112

```

```

alignment_block:
US-09-303-518D-465 x 069397 ..

```

```

Align seg 1/1 to: 069397 from: 1 to: 1455

```

```

64 GCACACGCTTCAGATTGGCAACGATTTCTTATCCGCGACGCTCTCGA 113
643 SerTyrAlaLeuLeuLeuThrAsnLysSerSerValProArgValTyrTy 659
114 CCCTCAGCATTTTGAACCCGACGGGGAATAC..... 144
659 rGlyAspMetPheThrAspAspGlyGlnTyrMetAlaHisLysThrIleA 676
145 .....CACCTATTC 153
676 snTyrGlnAlaIleGluThrLeuLeuLysAlaArgTyrLeuTyrValSer 692
154 GGCAGCAGGGGGGGAACCTTGGCCAGCGCAGCTCAT..... 189
693 GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGlnIleIleTh 709
190 .....ATGGATTTGGGAACATACAAAGCCATCGATTGGCGCAAC 229
709 rSerValArgTyrGlyLysGlyAlaLeuLysAlaThrAspThrGlyAspA 726
230 TGTTCATCCAGCAG.....GCGGCAATTAAGAAATATCGCTTAC 270
726 rGlnThrThrArgThrSerGlyValAlaValIleGlnGlyAsnAsnProSer 742
271 ATT...GTCCGCTTTCCGAT.....CA 290

```

743 LeuArgLeuLysAlaSerAspArgValValAlaLysMetGlyAlaAlaH 759
 291 CGGCGACGAGTCCATTCCTCCCTTC.....GACAC.....C 322
 759 slyAsnAlaLysAlaArgProLeuLeuThrThraspMetGlyLeu 776
 323 ATGCTCTACATTCGATTCGATGAGCCGCTAGTCCGTTGACGATTC 372
 776 ysalATyHisSeraspGlnAlaAlaGly..... 786
 373 AGCTTACCGCATTCATTTGGACGATACGAA.....CACATCCCGC 416
 787 ...LeuValArgTyThrAsnAspArgGlyLeuLeuIlePheThrAla 802
 417 CGAC.....GGCTATGACGGCGCACAGGGCGGCTATCC..... 453
 802 aspIleLysGlyTyThrAlaAsnProGlnValSerGlyTyLeuGlyAla 819
 454 .GCTCCCAAGGCGGACGAGGATATATACGATACGATAAAGCGCTT 501
 819 rPalProValGlyAla.....AlaAla 826
 502 GCCCAATATTCGGCTCCACCTGACCGACACGACGACCGACACG 551
 827 AspGlnAspValArgValAlaAlaSerThrAlaProSerThrAspGly 843
 552 GCTTGTGACCGCTTCCCAATACCGCTAGTATGTCACGACGAGTAG 601
 843 sSerVal.....HisGlnAsnAlaAlaLeuAspSerArgValMet 857
 602 GCACGATTCACACGCGCCACCGCATACGCGCCGACGTCGACAGATG 651
 857 heGlnGlyPheSerAsnPheGlnAlaPheAlaThrLysGlnGlyTy 873
 652 GGCATGCCCGCGACGCTTTCAC.....GGCAC 680
 874 ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluTrpGly 890
 681 TGCAGATATTCGTCACAAATCATCGCGCGGACGAGAAATTTGCGCG 730
 890 lThrAsp.....PheGlnMetAlaProGlnTyValSer 902
 731 CAGCGATGCGCTG.....CAGGGATATAAGCAGGCTCAACAT 771
 902 erThrAspGlySerPheLeuAspSerValIleGlnAsnGlyTyAlaPhe 918
 772 GCTGTTATCAGCGCTTGGCTGCTTTCCACGAAACACAGATGGCGG 821
 919 ThrAspArgTyAspLeuGly...IleSerLysProAsnLysTyGlyTh 934
 822 CATCAACGATTTGGCAGAT...ATGGCGCACTCAAAAGACTATGCCG 868
 934 rAlaAspAspLeuValLysAlaIleLysAlaLeuHisSerLysGlyLe 951
 869 CAGCCATTCGCGATGGCGATGCCAAACCCACGACGACACAGCAT 918
 951 yValMetAlaAspTrpValProAspGlnMetTyThrAlaLeuProGly 967
 919 GAAGCGCTCAGCAATATCTTACGGCAGTCAATCCCGTCAAGGATGG 968
 968 GluValAlaThr..... 971
 969 AGCTGTTCG...GGAATATACGCTTGGCGGACATCACGCGACATCT 1015
 972 .AlaThrArgValAspLysTyGlyThr.....Pro 982
 1016 TCACGCGTCCGACGATG...GGCGATCGCATTCGCGAAGGAAATCC 1062
 982 alAlaGlySerGlnIleLysAsnThrLeuTyValAlaAspLysSer 998
 1063 GCGGTCCGACGACAT.....TTTGGCGATGGCGC 1091
 999 SerGlyLysAspGlnAlaLysTyGlyAlaPheLeuGlnGluLe 1015

1092 ATACGCCAAATACCGCTCCCTTACCATCCGAAATATCCGTCAACT 1141
 1015 uGlnAlaLysTyTrpGluLeuPheAlaArgLysGlnIle..... 1028
 1142 TGGAGCAGGGTTACGGCAAGAAACATCATCCTCCGAAACCGTCCG 1191
 1029SerThrGlyValProMet 1034
 1192 TCACACGAAAGATGTGAAA...CTGCCAACAACACCGCACCGAAGAC 1238
 1035 AspProSerValLysIleLysGlnTrpSerAlaLysTyThrAsnGlyTh 1051
 1239 CAACAGCGCGTTGACGCGTAAAGCGTTTCCGAAATTTGAAAAAGCT 1288
 1051 rAsnIleLeuGlyArgGlyAlaGlyTy.....ValL 1062
 1289 AATACGATACGAGAAATTAATACCGCTGTACCAAGTGAATCTATAC 1338
 1062 euLysAspGlnAlaThrAsnThr.....TyThrSerLeuValSer 1075
 1339 GAACCGCTCTTAATCTAAGGTTCTGCGATCGGCTCAT.....TC 1382
 1076 AspAsnThrPheLeuProLysSerLeuValAsnProAsnHisGlyTh 1092
 1383 TTGGTCTAATACCTGCCAGATTCATACGCAAAATTTACCAAGGAGTA 1432
 1092 rSerSerValThrGlyLeuValPheAspGlyLys.....Gly 1105
 1433 GAATCAGATATATCCACCTTAAATTAATCTCTCTTCACACCGCTACA 1482
 1105 yValTyTyTrpSerThrSerGlyAsnGlnAlaLysAsnAlaPheIle 1121
 1483 AAGACCTAATATGATATTGGATTAATTTGGTAATGATGACTAA 1532
 1122 LeuGlyAsnAsnTrpTyTrpPheAspAsnGlyTyMetValThr.. 1137
 1533 AGTCCATCAAGAACTAAGTCAAGATTTGATG.....CATG 1573
 1138 .GlyAlaGlnSerIleAsnGlyAlaAsnTyTrpPheLeuSerAsnGly 1154
 1574 TTCATTTGCTTAA.....ACAGAGAGAGACACTTGATGG 1611
 1154 LeGlnLeuArgAsnAlaIleTyTrpAspAsnGlyAsnLysValLeuSer 1170
 1612 GCTAGTACGATGCTAAG 1629
 1171 TyrglyAsnAspGlyArg 1176
 seq_name: sp_invertebrate:Q9VRM6
 seq_documentation_block:
 ID Q9VRM6 PRELIMINARY; PRT; 1984 AA.
 AC Q9VRM6;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE CG5146 PROTEIN.
 GN CG5146.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,


```

183 CGGTCATATTCGGATTTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGT 232
      |||||   |||   |||   ::||||||   ::
170 ATGserProtaArgArgIArgIysSerArgserProserProtaArgArgIyse 186
      |||:::   ::   :::   :::   ::
233 TCATCCGACGACGGCGGCCATTAAAGGAAATTCGGCTTCATTTGTCGGCTTT 282
      |||:::   ::   :::   ::   ::   |||

```

```

186 rserProValAlrAglgLuArlysrYsrHsrIserProrAgh 203
283 TCCGATCAGCGGCACGAGTCATTCCC..... 311
      |||::|||
203 lAsrThrLysSerArgSerProSerProLaProglULtLySglULys 219
312 .....CTGCACCAACCATTGCCACATTGCGATTCTGATGAACCG 352
      |||::|||
220 SerProglUleuProglUProSerVal.....Ar 229
353 GTAGTCCCGTTGACGAGTTCAG..... 374
      |||::|||
229 GmELyAsPseSerSerValGLnGLuaLthrSerThrSerAsPILeuL 246
375 .....CCTTACCGCATTCATTGGACGGATACGACAACCATCC..... 413
      |||::|||
246 ySaLaProLysProglUProValProglUProLYsgLUProSerProglU 262
414 .....CGCCGACGGCATAGACGGCGGCACAGCGCG 442
263 LysAsnSerLysLysgLUsLysgLUsThraProArgSerAr 279
443 CGGCGTATCCCGCTCCCAAAGCGC.....GAGGATATA 477
279 gSerLysSerArgSerArgThrArgSerArgSerProSerHisThrArgp 296
478 TACAGCTACAGCATTAAGAAGCGT.....TGCCAAATATACCG 515
      |||::|||
296 roAtgaLgaLgHlSArGerSerArgSerArgSerTYrSerProArgArg 312
516 CCTCACACCTGACCGACA.....CCGACGACCGGACCAACGGGTGTGCG 559
313 ProSerProArgArgArgProSerProArgArgThrProProArg 329
560 A.....CGTTTCCACCATACCGGTACTATGCTGACGACGAAGATRAGC 603
329 GmeLrProProProArgHlsArArgSerArgSerProglYarArg 346
604 GACGAT..... 610
346 rgArArgSerSerAlaserLaserSerLysSerSerSerSerSer 362
611 .....TCAAAGCGGCACCCGATACACCCCGGACGTGCACAGATGGCG 654
363 SerArgSerArgSerProLysLysProProLysArgThr....Se 377
655 AATGCGCGCAAGCTTTCACGGCGACTGCACATATGCTCAAAAACATCAT 704
377 rSerProProArgLysThrArgArgLeuSer..ProSerAlaserProp 393
705 CCGCGCGCGCACGAGAAATTGTGCGCGCAGCGGATCCGTGCAAGGTATA 754
393 roArgrArgrArgrHlsArgrProSerSerProalArThrPro..... 405
755 GCAGAGGCTCAAAACATGTGCTTANAGCAGCGCTTGGSCTCGTTTGACC 804
406 .....ProPr 407
805 GAACAACAGATGCGCGCATCAGCATTTGGCAGATATGGCGCAACTCAA 854
407 oLysThrArg...HisSerProThrProglInLaserASArThrArgL 423
855 AGACT.....AMGCGCAGACGCATCCCGG..... 880
423 ysSerArgValSerValSerProglUArgThrsrGlyLysValThrLys 439
881 ....ATTGGCACTCCAACCAACCCCAATGCGCACAGGACATAGAACCGTTC 927
440 HisLysGLYThrgLUlYArGLgUserProSerProalabroLysProAr 456
928 AGCAATATCTTTAAGCGAGTATCCCGCTAAAGAGATTGGAGCTGTTCG 977
356 glySvalGIueLusSerGIueSerGIugLiuaSPrySGlySerLysMeTaLa 473

```

```

978 GGGAAATACGCTTGGCGCCGATCAGGCAC..... 1009
      ::::::::::::::::::::
473 laalaaspservalglnlglnarglntgrrarglgnasnlgln 489
1010 .....ATCCTGTCAAGCGGTGCAGATGGGC 1035
      ::::::::::::::::::::
490 SerSerSerSerSerGlySerSerSerThrSerLuspslunrProly 506
      ::::::::::::::::::::
1036 GAGATCGCATTCGCGAAGGGAAT.....CCGC 1064
      ::::::::::::::::::::
506 sargSerHisVallysasnlglnValglYargargarglLeuSerp 523
1065 CGTCAGGAGCAATTTGGCCGTCGCGATACGCCAATACCCGTCCTT 1114
      ::::::::::::::::::::
523 roserArgSerAlaSerPro.....SerProArgLysArgGlnLys 536
1115 ..ACCATTTCCGGAATATCCGTT..CAAACTTGGAGCAGCGCTTACGGC 1158
      ::::::::::::::::::::
537 GluthrSerProArgMetGlnMetGlySarGlyrpglInsSerProValTh 553
1159 AAG.....AAACATACCT..CCTCAACCGTCGCGCGTC 1193
      ::::::::::::::::::::
553 rlySerSerArgArgargargargSerProSerProProAlaArgArg. 569
1194 AAGGGAAGATGTGAACACTGCAACCAACGCCACCCGNAAGCAAG 1243
      ::::::::::::::::::::
570 ArgArgSerProSerProAlaProProProProProProProProPro 586
1244 TGCCGTTTGACGGTAAGGGTTCCGAATTTTGAAGAACGCTAAATAC 1293
      ::::::::::::::::::::
586 O.....ArgArgArgArgSerp 592
1294 GATAGAGAAATTAATACCGCTGTACCAAGTAATGCTATAGATGAACC 1343
      ::::::::::::::::::::
592 roThrProProArgArgArgArgThrProSerPro.Pro..... 604
1344 CGTCTTAATCTTAAGGTTCTGTGCGATCGGCTCATTTGTCTATAA 1393
      ::::::::::::::::::::
605 .....ProArg..... 606
1394 CTGCGAAGATTCATATACCAAAATATTACCAAGCAAGTAGATATAT 1443
      ::::::::::::::::::::
607 .....ArgArgSerProSerProArg.....ArgTyr 615
1444 ATCCGACCT.....AAAAATTAATCTCTCTCAGACCGCTACCAAAA 1485
      ::::::::::::::::::::
616 SerProProLleGlnArgArgLysSerProSerProPro...ProLys 630
seq_name: sp_human:09036
seq_documentation_block:
ID 09036 PRELIMINARY; PRT; 1275 AA.
AC 09036;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016091; BAA83717.1; -.
FT NON_TER
SQ SEQUENCE 1275 AA; 136869 MW; 45C2B2F85E98A6F6 CRC64;

```

alignment_scores:

```

Quality: 139.00 Length: 465
Ratio: 0.685 Gaps: 21
Percent Similarity: 43.656 Percent Identity: 24.946
alignment_block:
US-09-303-518d-465 x 09036 ..
Align seg 1/1 to: 09036 from: 1 to: 1275
12 CGCAAAATATCCCTTATTCGTCCATCTGACGACTGCTGCGCATGC 61
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
76 PROGLNGLuArgSerGluSerAspSerSerProAspSerLysAlaLysTh 92
   : || | |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|
92 rArgThrProLeuArgGlnArgSerArg.....SerGlySerS 105
112 GACCGTCACGATTTGCAACCCGCGGAAATACCACTATTCGAGGAG 161
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
105 erProLlual...AspSerLysSerArgLeuSerProArgArgSerArg 120
162 GGGGGAACCTTGC.....CGAGCGCACCGGTCATA 190
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
121 SerGlySerSerProGluValLysAspLysProArgAlaAlaProArgAl 137
191 TCGGATTTGGAAACATACAAAGCCATCACTTGGCAACCTGTCATCCAG 240
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
137 aglnSerGlySerAspSerSerPro.....GluProLysAlaProA 151
241 CA.....GGGCGCATTAAGGAATAT 263
   ||
151 laProArgAlaLeuProArgArgSerArgSerGlySerSerLysGly 167
264 CGGCTCATTTGTCGCTTTTCGATCCGCGCAGCAAGTCATTCGCCCT 313
   ||| ||| |||:::|:::|:::|:::|:::|:::|:::|:::|
168 ArgGlyProSerProGluLysSerSerThnGluSerSer...Progl 183
314 TCGAACAACATGCCCTCACATTCGATTCGATGAGCCGCTAGTCCCT 363
   ::||| ||| |||||:::|:::|:::|:::|:::|:::|:::|
183 whlProProLys.....SerArgThnAlaAlaArgA 193
364 GAGGATTCAGCCTTTACCGCATCCATTGGAGCGATACCAACCATCC 413
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
193 rglYSerArgSerSerProGluProLysThrLysSerArgThrProPro 209
414 CGCCGACGCGTATGACGGCCACACGGGCGG.....CGGCT 448
   ||||| |||:::|:::|:::|:::|:::|:::|:::|:::|
210 ArgArgArgSerSerArgSerSerProGluLeuThrArgLysAlaArgLe 226
449 ATCCCGCTCCCAAG.....CGCAGGGGATATA 477
226 userArgArgSerArgSerArgSerSerSerProGluThrArgSerArgT 243
478 TACAGCTACGACATTAAGGCGTTGCCCAAAATATCCGCTCAACCTGAC 527
   ||||| |||||:::|:::|:::|:::|:::|:::|:::|:::|
243 hrProProArgHlsArgArgSerProSerValSerSerProgluProAla 259
528 CGACAA.....CCGCAGCA 541
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
260 GLLuYSerArgSerSerArgArgArgSerArgSerArgSerProArgTh 276
542 CGGACACAGCGCTTGCA.....CCGTTCCACAAATACCGGTAGT 582
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
276 rLysThrThrSerArgArgGlyArgSerProSerProLysProArgGlyL 293
583 ATGCTGACGCAAGAGTAGG...CGACGATTCAAACGGCCACCGCG... 626
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
293 euglnArgSerArgSerArgSerArgArgLulysThrArgThrArg 309
627 ATACAGCCCGAGCTGACAGATCGGCAATGCGCCGCAAGCTTTAAAG 676
   ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
310 ArgArgAspArgSerGly.....SerSerGlnse 319

```



```

602 GCGAGGATTCGAAGCGCCAGCCGATACAGCCCGAGTCGAGCATCG 651
    ::::::::::::::::::::
657 hegluglypserasnpheglalphealatrlyslgslgltyr 873
    ::::::::::::::::::::
652 GGCATGCGCGCGAAGCTTTCAC.....GGCAG 680
    ::::::::::::::::::::
874 Thrasnvalallealalysasnvalasplysphaelaglturpglyva 890
    ::::::::::::::::::::
681 TGCGATATCTCTCAAAAACATCATCGCGCGCGAGGAGAAATTCGCGC 730
    ::::::::::::::::::::
890 Lthrasp.....Pheglumetalaproglmlyrvalisers 902
    ::::::::::::::::::::
721 CAGCGATGCGCGT.....CAGGATATACGAGGCTCAACATT 771
    ::::::::::::::::::::
902 ertthraspglyserpheleuaspservalileglnasnglytyralaphe 918
    ::::::::::::::::::::
772 GCTGTTATGACGGCTTGCTGCTCTTCACGCAAAAACAAGATGCGCG 821
    ::::::::::::::::::::
919 Thraspargtyraspleugly...ileserlyspasnlstyrglyth 934
    ::::::::::::::::::::
822 CATCAACGATTTGGCAGAT...ATGGCGCACATCAAGACATGCGCGAG 868
    ::::::::::::::::::::
934 ralaaspaspleuvallysalallelysalaleuhserslysllylel 951
    ::::::::::::::::::::
869 CAGCATCGCGATTTGGCGACGTCGCAAAACCCCAATGCCGCAAGGATA 918
    ::::::::::::::::::::
951 ysvalmetalaasprtyrpalproaspnlmetlyralaleuProglulys 967
    ::::::::::::::::::::
919 GAGCGCTGACGATATCTTTAGCGGACGTATCCCGCTCAAGGAGATTGG 968
    ::::::::::::::::::::
968 Gluvalvalthr..... 971
    ::::::::::::::::::::
969 AGCTTTCG...GGAATACGCGTGGGCGCATCAGCGCATCTGT 1015
    ::::::::::::::::::::
972 .Alathrtyrvalasplystyrglythr.....Prov 982
    ::::::::::::::::::::
1016 TCAGCGGCTCCGATG...GGCGAGATCGCATTCGCGAAGGAAATCC 1062
    ::::::::::::::::::::
982 alalaglyserglnlelysnthrleuvalvalaspglylysser 998
    ::::::::::::::::::::
1063 GCGGTCAGCGCAAT.....TTGGCGATGCGCGC 1091
    ::::::::::::::::::::
999 Serglylyaspnglnalalalystyrglyalalpheleugltule 1015
    ::::::::::::::::::::
1092 ATACCGCAATACCGCTCCCTTACCATTCGCGAATATCCGTTCAACT 1141
    ::::::::::::::::::::
1015 uglinalalystyrrprogluleuaphealarglyslntle..... 1028
    ::::::::::::::::::::
1142 TGGACGCGTTAGCGCAAAACATCACCCTCTCAACCGTCCGCGC 1191
    ::::::::::::::::::::
1029 .....Serthrglyvalpromet 1034
    ::::::::::::::::::::
1192 TCAGCGAAGAAATGTGAA...CTGGCAAAACAGCCGCAAGAC 1238
    ::::::::::::::::::::
1035 AspProseVallysllelysglnttrseralalystyrrpheasnglyth 1051
    ::::::::::::::::::::
1239 CAAAGTCCCGTTGACGGTAAAGGCTTCGAAATTTGAAAAAGCGTAA 1288
    ::::::::::::::::::::
1051 rsnlleuglyargglyalaglytyr.....ValL 1062
    ::::::::::::::::::::
1289 AATGATACGAGATTAATACCGCTGTACCAAGATGATCTATACAT 1338
    ::::::::::::::::::::
1063 eulysaspinalalthrasnthr.....Tyrrheserleuvalser 1075
    ::::::::::::::::::::
1339 GAACCGCTTAAATCTAAAGTTCGTGCGATCGGCTCAT.....TC 1382
    ::::::::::::::::::::
1076 AspAsnthrPheleuprolysserleuvalasnpProasnhlslythse 1092
    ::::::::::::::::::::
1383 TTGGCTATACCTGCCAGATTCATACGCAAAATTTACCAAGGAGGTA 1432
    ::::::::::::::::::::
1092 rserSerValthrglyleuvalPheaspglylys.....GlyT 1105
    ::::::::::::::::::::
1433 GAATCATGATATATCCACTATAAAATTTACTCTCTTCAAGACCGCTACA 1482
    ::::::::::::::::::::

```

```

1105 yValtyrtyrserthrserglyasnlnalalalysasnlabheleser 1121
    ::::::::::::::::::::
1483 AAGGACCTAATAATGATATTTGGATTAATTTGGTATGATGATGACATA 1532
    ::::::::::::::::::::
1122 LeuglyasnasntrpttyrtyrphesapnasnglytyrMetvalthr.. 1137
    ::::::::::::::::::::
1533 AGCTCATCAAGACTAAGTTCAGAAATTTGATG.....GATG 1573
    ::::::::::::::::::::
1138 .glyalaglnserlleasnnglyalalasnlytyrtyrPheleuserAsnglyI 1154
    ::::::::::::::::::::
1574 TTCATTTGCTTAA.....ACAGGAGAGACACTTGGATGG 1611
    ::::::::::::::::::::
1154 ledlnleuargasnalaletyrraspasnglyasnlysalvalleusertyr 1170
    ::::::::::::::::::::
1612 GCTAGTAGGATGTGTAAG 1629
    ::::::::::::::::::::
1171 Tyrglyasnaspolyarg 1176
    ::::::::::::::::::::

```

seq_name: sp_human:015038

seq_documentation_block:

```

ID O15038 PRELIMINARY; PRT; 1783 AA.
AC O15038;
DT 01-JAN-1998 (TRENBLER, 05, Created)
DR 01-MAY-2000 (TRENBLER, 13, Last sequence update)
DE 01-DEC-2001 (TRENBLER, 19, Last annotation update)
DE KIA0324 PROTEIN (FRAGMENT).
GN KIA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL, AB002322; BAA20782.2;
FT NON-TER 1
SQ SEQUENCE 1783 AA; 190940 MW; 660302F6PD4179AB CRC64;

```

alignment_scores:

```

Quality: 139.00 Length: 465
Ratio: 0.685 Gaps: 21
Percent Similarity: 43.656 Percent Identity: 24.946

```

alignment_block:

US-09-303-518D-465 x 015038 ..

```

Align seg 1/1 to: 015038 from: 1 to: 1783
12 CCGCAAAATATCCCTTATTCGTGCATACTGCGAGTGCCTCCGATGC 61
    ::::::::::::::::::::
584 ProglnglnargserluseraspserserProaspserslysalalysrh 600
    ::::::::::::::::::::
62 ATGCACACGCGCTCAGATTGGCAACGATTTTCGCGGAGGTTCTC 111
    ::::::::::::::::::::
600 rargthProleuarglnlArgserarg.....SerGlySers 613
    ::::::::::::::::::::
112 GACCGTCAGATTTGCAACCGGAGGAAATACCACTATTCGCGAGGAG 161
    ::::::::::::::::::::
613 ertProglulval...AspserlySerArgleuserProargdysersarg 628
    ::::::::::::::::::::
162 GGGGGAACCTTGC.....CGAGCGAGCGGTCATA 190
    ::::::::::::::::::::
629 SerGlyserSerProglulvallysalplysProarglalaalaproargal 645
    ::::::::::::::::::::

```

```

191 TCGATTGGAAACATCAACCATGTCGACCTGTTTCATCCAG 240
    ::::: ||||| ::::: |||||
645 aglnserglyseraspserPro.....gluprolysalaproA 659
241 CA.....GGCGCCATTAAAGAAATAT 263
    ||::|||
659 laProAlaLeuProAlaArgSerArgSerglySerSerserlysgL 675
264 CGGCTACATATTCGCTTTCCGACAGCGGACGACAAATCCATCCCT 313
    ||||| ||||| ::::: |||||
676 ArgglyProserProgluglySerSerserThrgluserSer...Progl 691
314 TCGAACACCATCCCTACATTCGATTCGATTCGATTCGATTCGATTC 363
    ::::: ||||| ||||| |||||
691 whlProProlys.....SerArgThrlaalaG 701
364 GACGATTCAGCTTTACCGCATTCATGGACGATAGCAACACCATCC 413
    ::::: ||||| ||||| |||||
701 rgllySerArgSerSerProgluprolysThrlYserArgThrProPro 717
414 CGCGACGCGCTATGACGGGCCACAGGCGG.....CGGCT 448
    ||||| ||||| ::::: |||||
718 ArgArgArgSerSerArgSerSerProgluLeuthrArglyAlaArgLe 734
449 ATCCGCGCTCCCAAGG.....CGCGAGGATATA 477
    ||||| |||||
734 userArgArgSerSerAlaSerSerSerProgluThrArgSerArgT 751
478 TACACCTAGCATTAAGCGGCTTCCCAAAATATCCGCTCAACCTAC 527
    ||||| ||||| |||||
751 hrProProAlaThrlArgSerProSerValSerSerProgluProAla 767
528 CGACAA.....CGCGACCA 541
    ::::: |||||
768 GlulysSerArgSerSerArgArgArgSerSerAlaSerSerProAla 784
542 CCGGACACGCGCTGTGCA.....CGCTTCCACATATCCGCTAGT 582
    ::::: ||||| ||||| |||||
784 rlySerThrlSerArgArglyArgSerSerProlysProArglyL 801
583 ATGCTGACGCAAGAGTAGG...CGAGGATTCAAACCGCGCACCGC... 626
    ::::: ||||| ||||| |||||
801 euglnArgSerArgSerSerArgArglyglulysThrlArgThrlArg 817
627 ATACAGCCCGAGCTGACAGATCGGCAATCCGCGCAAGCTTTCAACG 676
    ::::: ||||| |||||
818 ArgArgAspArgSergly.....Serserglnse 827
677 GCACGCGCATATGCTCAAAAACATCATCGCGCGGACGAGAAATGTC 726
    ||||| ||||| ::::: |||||
827 rThSerArgArglyglulysSerArgSerArgSerArgValThrlArg 844
727 GCGGACGCGCATGCGCTGACAGGTATAAGCAAGGCTCAACATTCGT 776
    ||||| |||||
844 rArgArgly..... 847
777 TATGACGCGCTGGTCTGCTTCCACCGAAACAGATGCGCGCATCA 826
    ||||| ||||| ::::: |||||
848 .....GlyserglyThrlSerArgSerPro...AlaArggl 859
827 ACGATTTGGCAGATATGGCGCACTCAAGACTATGCCGACGACGCGCATC 876
    ::::: ||||| |||||
859 ngluserSerArgThrlSerSerArgArgArglyArgSerArgThrlP 876
877 CGCGATTGGCAGTCACAAACCCCAATGCCGACACAG..... 914
    ||||| ||||| ::::: |||||
876 roProThrlSerArglyArgSerArgSerArgThrlSerProAlaProTIP 892
915 CATGAACCGCTCAGCAATATCTTTACGCGAGTCAT..... 950
    ||||| ||||| ::::: |||||
893 lylArgSerArgSerArgAlaSerProAlaThrlThrlArgArgSerArgse 909

```

```

951 .....CCCCGTCAAAAGGATTGAGCTGTTCGGGCAAA 984
    ||||| ||||| |||||
909 rArgThrProLeuIleSerArgArgSerArgSerArgThrlSerProV 926
985 TACGCGCTTGGCGCATCAGCGACATCTGTGACAGCGCGCAGATGGG 1034
    ::::: ||||| ||||| |||||
926 alSerArgArgArg...SerArgSerArgThrlSerValThrlArgArg 942
1035 CGAGATCGCATTCG.....CGAAAGGAAAT 1060
    ::|||::|||
942 erArgSerArgAlaSerProValSerArgArgArgSerArgSerArg 958
1061 CCGCGCTTCAGCAATTTGCGATCGCGCATACGCAATATCCGCTC 1110
    ||||| ||||| ::::: |||||
959 ProProValThrlArgArgArgSerArgSerArgThrlThrlArgArg 975
1111 CTTTACCATTCGCCGAATATCCGTTCAACCTGAGACGCGTACGCA 1160
975 g.....ArgSerA 978
1161 AGAAACATCACCCTCCTCAACCGTCCGCTCAACGCAAGCA 1204
    ::::: ||||| |||||
978 rgSerArgThrlProProValThrlArgArgArgSerArgSerArg 992

```

seq_name: sp_human:060382

seq_documentation_block:

```

ID 060382 PRELIMINARY; PRT: 1791 AA.
AC 060382;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KIA00324 (FRAGMENT).
GN KIA00324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBT_TaxID=9606;
ON 111
RP SEQUENCE FROM N.A.
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,
RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.,
RT "Sequencing of Human Chromosome 16p13.3."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
RT Comparison Analysis (SCAN) System."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC004493; AAC08453.1; -.
FT NON_TER
SQ SEQUENCE 1791 AA; 191306 MW; 3A7B5530AE95F3E CRC64;

```

alignment_scores:

Quality:	139.00	Length:	465
Ratio:	0.685	Gaps:	21
Percent Similarity:	43.656	Percent Identity:	24.946

alignment_block:

US-09-303-518D-465 x 060382 ..

Align seg 1/1 to: 060382 from: 1 to: 1791

```

12 CCGCAAAATATCCCTATTCCTGTCATCTGCGACGATGCTCGCGAGC 61
    |||||::|||
585 PROGLUGLArgSercluserAspserProAspserlyAlaLysTh 601
    ||||| ||||| ::::: |||||
62 ATGACACCGCTCAGATTGCGCAACGATCTTTATCCGCGAGGTTCTC 111
    ::||| |||||

```


RA Laako P, Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Moberry C., Morris J., Mostreil A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.,
RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-T., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhu S., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AEO03591; AAF51592.1; -
DR HSSP: P01180; 2BN2.
DR FlyBase: FBgn0036985; CG5847.
DR InterPro: IPR001507; zone_pellucida.
DR SMART: SM00241; zp_1
SEQUENCE 2284 AA; 248295 MW; F495EAB8CB844674A CRC64;

```
alignment_scores:
  Quality: 139.00
  Ratio: 0.728
  Percent Similarity: 40.812
  Length: 468
  Gaps: 24
  Percent Identity: 23.932
```

```
alignment_block:
```

US-09-303-518D-465 x Q9VPG1 ..

Align seg 1/1 to: Q9VPG1 from: 1 to: 2284

```

2420 TGGTTCATCCACAGACGGCCGCAATTAAGAAATCGCATATTGCGCC 279
1022 CysGlnProAlaThrArgPro..... 1022
280 TTTTCCGATCAGCGGACAGAG..... 301
1029 ...ProIleThrThrSerIleProArgCysTyrProGlySerLeuAspP 1044
302 ..TCGATTCCCGCTTGACACACCATGCGTCACACTTCCGATTCTGATGAG 349
1044 roGluCysGlnProSerThrTyrLeuProProThrProVal.....Arg 1056
350 CCGGTAGTCCCGTTGACGAGTTCAGCGTTTACCGCATCCATTGGAGCGA 399
1059 ThrThrValPro..... 1062
400 TAGACACCATCCCGCGCGCGCTATGACGGGCCACAGGGCGCGCGCTA 449
1063 .....ThIleProThr.....ArgI 1063
450 TCCCGCTCCCAAGGGCGCA..GGGATATATACAGCTACGACATPAAAG 496
1069 leProValThrThrSerIlysProAsnCysTyrProGlySerThrAspar 1085
497 GCGTTGCCCAAAATATCCG.....CCTAACCTGACGACACACCGC 537
1085 gArgCysProIysGlnProValThrThrProIysProArgCysTyrProg 1102
538 AGCAGCGGACAGCGCTTTCGACCGCTTCCACATACCGGAGTAGATCCT 587
1102 LyserProAsnProGluCys.....GlnIysAlaThrTyrSer 1114
588 GACGACAGAGATGAGCGGACGAGATTCAACGCG..CCAGCGATACAGCCCC 636
1115 ProProThrThrArgThrProValThrThrSerIlysProAsnCysTyrPr 1131
637 GAGCTGACGAGATCGGGCAATCGCGCGGAGGATTCAACGGCATCGGAGA 686
I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I

```

[illegible]

seq_name: sp_virus:090P86

seq_documentation_block:

ID 090P86 PRELIMINARY; PRT; 851 AA.

AC 090P86;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE MAJOR OUTER ENVELOPE GLYCOPROTEIN GP350.

OS Human herpesvirus 4 (Epstein-Barr virus).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus.

NCBI_TaxID=10376;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE-98203772; PubMed-9542645;

RA Lee W.K., Kim S.M., Sim Y.S., Cho S.G., Park S.H., Kim C.W.,

RA Park J.G.;

RT "B-lymphoblastoid cell lines from cancer patients";

RL In Vitro Cell. Dev. Biol. Anim. 34:97-100(1998).

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE-99072166; PubMed-9856346;

RA Chang S.H., Kim S.H., Lee W.K., Kim H.J., Choi S.H., Park J.H.,

RA Jang H.S., Chung G.H., Kwon T.H., Kim D.H., Yang M.S., Jang Y.S.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF172333; AAD51698.1; -

DR InterPro: IPR002965; P_Rich_extensn.

DR PRINTS: PR01217; PRICHEXTENS.

KM Envelope protein.

SQ SEQUENCE 851 AA; 89112 MW; 3B2C4862C6396DC8 CRC64;

alignment_scores:

Quality: 138.00 Length: 575

Ratio: 0.561 Gaps: 25

Percent Similarity: 42.783 Percent Identity: 21.739

alignment_block:

US-09-303-518D-465 x 090P86 ..

Align seg 1/1 to: 090P86 from: 1 to: 851

```

110 TCGACCGTCAGCATTTGCAACCGGAGGAAATACACATTCG.... 154
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 AsnThrThrAspIleThrValGlyAspAsnIleThrTyrSerValPr 334
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 .....GCAGCAGGGGGAACTTCCCGACGCGCATCATGCG 194
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
334 ometValThrSerGluAspAlaAsnSerProAsnValThrValThrAlaP 351
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
195 ATTTG.....GAACATCAACAGCCATGAGTGGGCA 226
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 heThrPAlaTrpAsnAsnThrGluThrAspPheLysCysAlaTyrThr 367
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
227 ACCTGTCATTCAGCAGGCGGCGCATTAAGAAATATGCGTACATTTC 276
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 LeuThrSerGlyThrProSerGlyCysGluAsnIleSerGlyAlaPheAl 384
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
277 CGCT.....TTTCGATCAGGCG..... 295
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
384 aserAsnArgThrPheAspIleThrValSerGlyLeuGlyThrAlaPhe 401
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 .....ACGAAGTCCATTCCCGCTTGACACACCATGCGTAC 331
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

401 etThrLeuIleIleThrArgThrAlaThrAsnAlaThrThrThrHis 417
332 ATTCGGATTCGTATGAGACCGGTAGTCCCGTTGACGGATTCAGCTTAC 381
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 LysValIlePheSerLys.....AlaProGluSerThrThrThr 430
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
382 CCGATTCATTTGGGAGGATGACAGACACATCCCGCGGCGGTATGACGG 431
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
430 rSerProThrLeuAsnThrThrGlyPheAlaAlaProAsnThrThrThr 447
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
432 GCCACAGGGCGGGCGGTATCCCGCTCCCAAGGCGCGGAGATATATCA 481
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 Ly.....LeuProSerSerThrHisValProThr 456
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
482 GCTACGACATTAAGCGCTG.....CCCAATATTCGCGCTCA.... 520
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 AsnLeuThrAlaProAlaSerThrGlyProThrValSerThrAlaAspVa 473
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 .ACGTGACCGACACCGGACGACGACGCGCTTTCGACCGTTTCCA 569
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
473 ThrSerProThrProAlaGlyThrThrSerGlyAlaSerProValThrP 490
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
570 CAATACCGGTAGTATGCTGACGCAAGAGTAGCGGCGGATCAACGCG 619
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 rSerProSer.....ProAlaGpAsnGlyThrGluSerLysAla 503
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
620 CCACCGGATACAGCCCGGAGCTGG..... 643
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
504 ProAspMetThrSerProThrSerAlaValThrThrProThrProAsnAl 520
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
644 .....ACAGATGGGCGCATGCGCG 662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
520 arThrSerProThrProAlaValThrThrProThrProAsnAlaThrSerP 537
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 CGAAGCTTTCACGCGGCGCATGATGCTCAAAACA..... 700
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 roThrLeuGlyThrSerProThrSerAlaValThrThrProThrPro 553
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 .....TCATCGGCGGCGGAGGAAATGTCGGCGGACGATGCGCG 744
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
554 AsnAlaThrSerProThrValGlyGluThrSerProGluAlaAsnThrTh 570
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
745 CAGGATATAGCAGAGCTCAACATTCG.....CTGTATCGACGCGT 788
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
570 rAsn.....HisThrLeuGlyGlyThrSerThrProThrPro 582
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
789 GGGTCTGCTTTCACGCAAAACAGATGCGCGGATCAACGATTTGGCG 838
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
582 aValThrSerProThrLysAsnAlaThrSerAlaValThrThrGlyGln 598
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
839 .....ATATGGCGCAACTCAAGAGCTATGCGGACGACGATCGCG 879
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
599 HisAsnIleThrSerSerThrThrSerMetSerLeuArgProSerG 615
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
880 GATTGGGACATCCAAACCCCAATGCGGCAAGCATGAAGCCGCTGAC 929
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
615 rIleSerGluThrLeuSerProSerThrSerAspAsnSerThrSerHis 632
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
930 CAATATCTTACGCACTATCCCGCTCAAGGATGAGCTGTTCGCG 979
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
632 etProLeuLeuThrSerAlaHisProThrGlyGlyGluAsnIleThrGln 648
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
980 GAAATATGCGCTTGGCGGCGATCAGGCGACATCTGTGCAAGGCTGCG 1029
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
649 ValThrProAlaSerThrThrHisHisValSerThrSer..... 662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1030 ATGGGCGAGATCGCATTTGCCGAAGGAAATCCCGCGTACGCAACATTT 1079
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
662 ..... 662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1080 TGCGGATGCGGCGATACGCAATACCGCTCCCTTACATTCGCGCAATA 1129
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 .....SerProAlaProArgProGlyThrThrCysGlnAla 674
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

CC      -1-  CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS
CC      549 ONWARDS DUE TO A FRAMESHIFT.
DR      EMBL; AL035077; CA22867.1; -.
DR      EMBL; D89218; BA013879.1; AUT_FRAM.
KW      Hypothetical protein.
FT      DOMAIN 368 373
FT      COORDINATE 433 433
SO      SEQUENCE 615 AA; 65329 MW; E0D4C95666775A3 CRC64;

Alignment_scores:
    Quality: 137.50      Length: 572
    Ratio: 0.476      Gaps: 32
Percent Similarity: 50.524      Percent Identity: 21.503

Alignment_block:
US-09-303-518D-465 x 094497 ..

Align seg 1/1 to: 094497 from: 1 to: 615

43 GCAGTGCTCCCTGCCGATGCATGCACACGCCCTCAGATTGTGGCAACGATTTC 92
|||||
:::
16 AAlaValProLysSerLeuAlaGlySerThrSerAsnIleProAsnGluLeu 32
|||||
:::
93 TTTATATCCGCGAGGTTCTCGACCGTCAGCATTTGCAACCCGACGGAAT 142
|||||
:::
32 upheVal.....GlyProAspGlyArgP 40
|||||
:::
143 ACCACGATATTCGGCAGCAGGCGGAGCACTCCGAGCGCAGCAGCGTCATATC 192
|||||
:::
40 roLeuSerGlnAsnAlaGlnGlyLaserLysSerSerLysValVal 56
|||||
:::
193 GGATTGGGAACATACAAAGCCATCACTGGGCAACCTGTTCATCCAGCA 242
|||||
:::
57 ProGlnAlaLeuValProGluAspAspIle..... 66
|||||
:::
243 GCGCGCATTTAAAGAAATATCGGCTAC.....ATTGCCGCT 280
|||||
:::
67 ArgAlaValGlnGlyIleLeuAsnTyrGlyGlySerAsnAspAsnArgP 83
|||||
:::
281 TTTCCGATCACGGGCAC.....GAACTGCATTTCCCTTCGACAAC 321
|||||
:::
83 roValSerHisThrHisThrPheValGluLeuGlnLysSerHisGlnAsn 99
|||||
:::
322 CATCCGTCACATTCGCGATCT.....GATGAAC 350
|||||
:::
100 HisLeuThrGluAsnAspArgAsnPheGlyThrSerArgLysAspAsp 116
|||||
:::
351 CGTAGTCCGCTTGACGATTCAGCCTTTACCCGATCATTTGGACGAGAT 400
|||||
:::
116 LAlaProAsnAlaAspGlyValArgArgLeuArg.ThrSerGlySerSer 132
|||||
:::
401 ACG.....AACACATCCCGCGGAGCGGCTATGACGGGCACAGGGCGC 444
|||||
:::
133 ThrIleuSerAsnAlaProProSerAlaAsnValSerLysAlaSerSer 149
|||||
:::
445 GGCATATCCCGCTCCCAAGGCGGAGGATATATATACAGCTACGACATAA 494
|||||
:::
149 AsnLeuSerLeu.....AlaSerLeuAlaLysThrGlnProGluA 163
|||||
:::
495 AGGCGTTGCCAAATATCCGCTCAACTGACCGGCAACCGCAGCA 541
|||||
:::
163 TgaLAtThrProGluValLysValProLeuAsnProAspThrGlySerVal 179
|||||
:::
542 .....CCGACAAAGGCTTGC..... 558
|||||
:::
180 ProLeuIleHisProGluGlnThrAspArgGlyLeuProTyrAlaProAs 196
|||||
:::
559 GAACGTTTCCACAAATACCGGATGATG...CTGAACGCAAGGATAGGC. 603
|||||
:::
196 RgIlyuSerPheHisAsnSerGlySerLeuLysLeuProLysGlyLaserL 213
|||||
:::
604 GAGCGATTCGAACGCGGCAACCGGATTAAGCGGCTACAGCGGCTAC

```

```

213 eugluaspLeuSerArg.....SerProSerArgAlaValLeu 225
      ::::: ||| |||||::: |||
643 GACAGATCGGGCAATGCGCCGGAAGT..... 669
      ::::: |||||::: |||||
226 AsnGluaspGlyAsnValaspGluCysAlaProProGluProTrpGluAs 242
      ::::: |||||::: |||||
670 ....TTCAACGGCAGCTGACATATCGTCAAAAC...ATATGGC... 708
      ::::: ||| |||||::: |||||
242 nglutyrAsnGluValLeuaspAspValGluAsnAlaValValGlyThrS 259
      ::::: |||||::: |||||
709 .....CGCGCAGAGAAATGTGCGC 729
      ::::: |||||::: |||||
259 erProLeuGluTyrThrSerLysProLeuAlaAlaAsnArgGlnArgSer 275
      ::::: |||||::: |||||
730 GCAGCGCATGCCGTGCAGGGTATAGCGAGCGCTCAACATGTGCTTAT 779
      ::::: |||||::: |||||
276 ThrAlaAsp.....LeuThrGluSerAspAsnIleCys.... 286
      ::::: |||||::: |||||
780 GCAGCGCTTGGTCTGCTTCCACCGAAACACAGATGCGCGCATCAAG 829
      ::::: |||||::: |||||
287 .....GlyLeu.....ThrAlaGlyLysSerAspProValThr 298
      ::::: |||||::: |||||
830 ATTGGCAGATATGCGCACTCAAGACTATGCCGACGACCGATCGC 879
      ::::: |||||::: |||||
298 spVal...AspGluSerGlnThrIleAspGlnSerIle..... 310
      ::::: |||||::: |||||
880 GATTGGGAGTCGCAAAACCCCAATGCCGACAGCATAGACCGCTGAC 929
      ::::: |||||::: |||||
311 .....ProGluAlaGluLysGlyPheTyrThrLysAs 321
      ::::: |||||::: |||||
930 CAATATCTTACGCGATCATCCCGTCAAGAGGATGGAGCTGTGCGG 979
      ::::: |||||::: |||||
321 pGlyGluGlyThrAlaGlyLeuProPheAsp.....IleValS 334
      ::::: |||||::: |||||
980 GAAATATACGGCTGGCGGCATCAGCGACATCTGTACACGGTGGCAG 1029
      ::::: |||||::: |||||
334 erAsnLeuAspIleProAsnGluAsnAlaHisGluSerSerArgSerLys 350
      ::::: |||||::: |||||
1030 ATGGCCGAGATCGCATGGCCGAAAGGAAATCCCGCCGACGCAATTT 1079
      ::::: |||||::: |||||
351 LysLysSHStHngly...ProSerLeuSerSerAlaSerGlnProSerAl 366
      ::::: |||||::: |||||
1080 TGCCGATCGCGCATGCGCAATATCCCGCTTACCATTCGCGAAT. 1128
      ::::: |||||::: |||||
366 alaSerSerSerSerSerSerGluProSerAsnLeuAspLysIleAsn 383
      ::::: |||||::: |||||
1129 ..ATCCGTTCAAACTGGAG.....CAGCGTTAC 1155
      ::::: |||||::: |||||
383 spValLysLysAsnIleGluValSerAlaAsnGluProGlnProArgPro 399
      ::::: |||||::: |||||
1156 GGCAAGCAAAACATCATCCCTCA.....ACCGT 1184
      ::::: |||||::: |||||
400 ValLysGluAspValProLysSerGlnValGlyGlyGluThrAspThrTh 416
      ::::: |||||::: |||||
1185 GCCCGCGTCAACGGAAGAATGTGAACGTGGCAAAACAAACGACCGCA 1234
      ::::: |||||::: |||||
416 tAspValIleAsnAsnSerThrProLysGluGluThrGluGluSerPro 433
      ::::: |||||::: |||||
1235 AGACCAAACTGCCGTTTGACGTGAAGGTTTCCGAATTTTGAAAAAGAC 1284
      ::::: |||||::: |||||
433 erThrGluLeuProGluThrGlyLysGluGlnProProAsnLysAlaGlu 449
      ::::: |||||::: |||||
1285 GTAATAATACGATAGCAATTAATACCGTGTACCA...CAAGTAATCC 1331
      ::::: |||||::: |||||
450 ProAlaValAlaProThrGluAlaSerThrLysProSerGluAlaAlaG1 466
      ::::: |||||::: |||||
1332 TATAGATGAACCGCTCTTAAT.....CTTAAGCTTCTGTGCGATCGG 1375
      ::::: |||||::: |||||
466 uGluSerThrProArgPheSerValArgProAsnLysPheThrGlySer 483
      ::::: |||||::: |||||
1376 CTAATCTTGG.....TCTATACTGCGCAAGTATCATACGCAAAATTA 1419
      ::::: |||||::: |||||

```

```

483 rGAlaGlyPheValAlaAlaLeuGluSerArgLeuGlnLysGlyProLeu 499
      ::::: |||||::: |||||
1420 CCAAGGCAAGGTAGATCATCATCCACTTAATAATTCCTCCCTC 1469
      ::::: |||||::: |||||
500 MetArgSer.....PheValProAsnLysSerLysSerProSe 512
      ::::: |||||::: |||||
1470 AGCACCCTACCCAAAGACACTAATATGATATTTGGATAATTTGGTA 1519
      ::::: |||||::: |||||
512 rGly.....ThrLysSerProAlaSerGlyGluThrSerGluAlaGly 527
      ::::: |||||::: |||||
1520 ATGATGACTTAA 1533
      ::::: |||||::: |||||
527 alLysGlnThrGlu 531
      ::::: |||||::: |||||
seq_name: sp_Invertebrate:Q9V515
seq_documentation_block:
ID Q9V515 PRELIMINARY; PRT; 746 AA.
AC Q9V515;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG8181 PROTEIN.
GN CG8181.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington I., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003835; AAF59007.1;
DR FlyBase: FBgn0033361; CG8181.

```


DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SO SEQUENCE 746 AA; 78593 MW; FB6F9F8DA3027334 CRC64;

alignment_scores:
 Quality: 137.00 Length: 455
 Ratio: 0.745 Gaps: 19
 Percent Similarity: 40.440 Percent Identity: 20.879

alignment_block:

US-09-303-518d-465 x Q9V515 ..

Align seg 1/1 to: Q9V515 from: 1 to: 746

```

110 TCGACCGTCAGCATTTGACACCGGAGGAATACACCTATTGCGGACG 159
    ::::::::::: :::: ||||| :::::::::::
244 ThrThrThrThrAlaAlaProThrThrThrThrThrThrAlaAla 260
    ::::::::::: :::: ||||| :::::::::::
160 AGGGGGGAACTTCCGAGCGGAGCGGTC..ATATCGATTGGGAACAT 206
    :::: ||||| :::: |||||
260 rThrThrThrThrProThrProThrLeuAlaThrThrThrSerLysProH 277
    :::: ||||| :::: |||||
207 ACAAGCCATCATGTTGGCAACCTGTTATCCAGCAGCGCCATTA... 253
    :::: ||||| :::: |||||
277 LysThr.....HisSerSerArgArgProAlaPro 287
    :::: ||||| :::: |||||
254 .....AAGAAATATCGGCTACATTTGCGCTTTCGATCAGCGGAC 297
    :::: ||||| :::: |||||
288 AlalysLysProValSerThrThrThrThr.....AlaSerGlyThr 302
    :::: ||||| :::: |||||
298 GAGTCCATT..... 307
    :::: ||||| :::: |||||
302 rGlnSerThrArgLysProAlaAspSerLeuAlaSerGlySerThrThra 319
    :::: ||||| :::: |||||
308 .....CCCCCTTCGACACCATGCGCTCCATCCGATTCTGATAG 349
    :::: ||||| :::: |||||
319 laaenglyThrThrAlaThrThrLeuPro.....AlaArgArg 331
    :::: ||||| :::: |||||
350 CCGGTAGTCCGTTGACGAGTATGACCTTTACCGCATTCATTGGAGCA 399
    :::: ||||| :::: |||||
332 ProValThrGlnThrSerSerSerSerThrThrThrThrThrSe 348
    :::: ||||| :::: |||||
400 TACGACACCC.....ATCCGCCGACGGCTATGA 428
    :::: ||||| :::: |||||
348 rThrThrThrThrValPheLysAspGluLeuProLysAsnArgThr 365
    :::: ||||| :::: |||||
429 CGGGCCAC.....AGGGCGGGGCTATCCGGCTCCCAAG 463
    :::: ||||| :::: |||||
365 hrThrArgLeuProGluArgThrThrThrAlaProArgProArgProLys 381
    :::: ||||| :::: |||||
464 GCGCGAGGATATATACAGCTACGACATAAAGCGCTTGCCCAAAATATC 513
    :::: ||||| :::: |||||
382 ProThrGlyLeuLeuVal.....LysValProAlaLeuLe 393
    :::: ||||| :::: |||||
514 CGGCTCAACCTGACGACACCGGAGCAGCAGGACGCTTGTCAGC. 562
    :::: ||||| :::: |||||
393 uAlaGlu.....ProThrThrThrThrValThrLysGlySerSerS 408
    :::: ||||| :::: |||||
563 .....GTTTCACAAATACCGGTAGTA 583
    :::: ||||| :::: |||||
408 erSerSerSerSerProLysProProAlaSerSerThrThrThrProLeuVal 424
    :::: ||||| :::: |||||
584 TGCTGACGCAAGAGTAGCGAGCGAGATTCAACGCCCA..... 622
    :::: ||||| :::: |||||
425 ThrLysLysLys.....ProThrThrThrThrAsnAlaProThrThrThr 439
    :::: ||||| :::: |||||
623 .....CCGATTAAGCCCGGAGCGTGCAGACATC 650
    :::: ||||| :::: |||||
439 rThrThrThrThrThrThrThrThrThrThrProLysProThrArg 456
    :::: ||||| :::: |||||
651 GGGCAATCGCGCGCAAGCTTTCACAGCGCACTGAGATATGTCAAAAACA 700
    :::: ||||| :::: |||||

```

```

456 rGlnThrLysProProThrThrThrThrThrThrThrThrLysAlaThrThr 472
    ::::::::::: :::: ||||| :::::::::::
701 TCATGGCGCGCGCAGAGAAATTCGCGCGCAGCGATGCCGTCAGGGT 750
    ::::::::::: :::: ||||| :::::::::::
473 ThrThrThrLysAlaThrThrThrThrThrThrThrAlaLysPro..... 486
    ::::::::::: :::: ||||| :::::::::::
751 ATAAAGCAAGGCTCAAAACATTGCTTATGCACGGCTGGCTGCTTC 800
    ::::::::::: :::: ||||| :::::::::::
487 .....LeuThrThrThrGluProProThrSera 496
    ::::::::::: :::: ||||| :::::::::::
801 CACCGAAACAGATGCGCGCATCAACGATTGGCAGATATGCGGCAC 850
    :::: ||||| :::: |||||
496 laProLeuThrThrThrThrThrLysThrLysThrLysLeuThrThrThr 512
    :::: ||||| :::: |||||
851 TCAAGACTATGCCGACGACGACCATCCGCGATTGGCAGTCCAAACCC 900
    :::: ||||| :::: |||||
513 ValGluValGluProProThrAsnAlaSerLysSerAsnAspThrLys 529
    :::: ||||| :::: |||||
901 AATGCCGACACAGCATAGAACCGCTCAGCATATCTTACGGCATCAT 950
    :::: ||||| :::: |||||
529 AlaPro..... 531
    :::: ||||| :::: |||||
951 CCCGTCMAAGGATTTGAGCTGTTGGGGAATAACGCTTGGCGGCA 1000
    :::: ||||| :::: |||||
532 ..ProSerAsnThrLeuProLeuThrThrGluSerSerGluThrAlaAsp 547
    :::: ||||| :::: |||||
1001 TCACGGCACATCTCTGTCMAAGCGGTCCAGATGGCGAGATCCGATTGCCG 1050
    :::: ||||| :::: |||||
548 LysThrSerAlaSerThrAla..... 554
    :::: ||||| :::: |||||
1051 AAGGGAATATCCGCGTCAGCGCATTTTGGCGATGCGCATACGCCAA 1100
    :::: ||||| :::: |||||
555 .....ProProSerAlaThr..... 559
    :::: ||||| :::: |||||
1101 ATACCGCTCCCTTACCATTCCCGAAATATCCGTTCAACTGGAGCACG 1150
    :::: ||||| :::: |||||
560 .....AspLysValValLysValAlaThrSer 569
    :::: ||||| :::: |||||
1151 GTTACG.....GCAAGAAAAATC 1170
    :::: ||||| :::: |||||
570 ValSerGluValGluThrHisGlyProGlyLysProGluLysThrThrLys 586
    :::: ||||| :::: |||||
1171 ACCTCTCAACCGTGGCGCGCTCAACGGAAGAAATGTGAACGTGGCA 1220
    :::: ||||| :::: |||||
586 sProProLysPro..... 590
    :::: ||||| :::: |||||
1221 CAACGCCACCGCAACACCAAGTGCCTTACGTTAAGGTTTCCGA 1270
    :::: ||||| :::: |||||
591 ..SerSerThrAlaLysProThrThrThrThrThrAlaAlaAlaThrThr 606
    :::: ||||| :::: |||||
1271 ATTTGAAAAAGACG 1285
    :::: ||||| :::: |||||
607 lIleValLysThrThr 611
    :::: ||||| :::: |||||

```

seq_name: sp_rodent:063625

seq_documentation_block:

ID 063625 PRELIMINARY; PRT: 1473 AA.

AC 063625:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE RA9.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

ON NCB1:taxid=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HIPPOCAMPUS;

RX MEDLINE=96293459; PubMed=8692929;

RA Yuxey A., Patudrajan M., Litlingtung Y., Joshi R.V., Gentile C.,

Gedara M., Corden J.L.;

RT "The C-terminal domain of the largest subunit of RNA polymerase II
interacts with a novel set of serine/arginine-rich proteins.";
RT Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).
DR EMBL: U49057; AAC52658.1;
DR SPROUCE 1473 AA; 161204 MW; 949EB6F5873989BF CRC64;

alignment_scores:

Quality: 137.00 Length: 624
Ratio: 0.566 Gaps: 33
Percent Similarity: 38.782 Percent Identity: 21.314

alignment_block:

US-09-303-518D-465 x Q63625 ..

Align seg 1/1 to: Q63625 from: 1 to: 1473

```

138 GAAATACCACTATTTCGCGACGAGGGGAACTTCGCGAGCGCGGTC 187
||||| ||| :||| |||||:||||| |||||
719 GUILLEIUPROTHGLIIEGLNGLYSERIALARGALAGLARG.. 734
188 ATATCGATTGGGAAACATACAAAGCCATCGTTGGCAACCTGTTATC 237
||||| :||| :|||
735 ..ProserProProAspProTTPAspa 743
238 CAGCAGGGGGCCATTAAAGAAATATCGGTACATTGT..CCGCTTTC 284
||| :||| :||| |||||:||||
743 spGLAspGLy.....ValSerCysThrProPhe 753
285 CGATCAGCGGCGACGAGTCCATCCCTTCGACAAACCATCCCTCAGATT 334
||||| :||| :|||
754 GLYSERGLUARGTHR.....ValThrCysValThr.. 764
335 CCGATTTCGATGAGCCGCTAGTCCCTTGACGATTTCAGCCTTACCGC 384
764 .....
385 ATTCATTGGGACGATACGAAACACCATCCCGCGAGCGTATGAGGGCC 434
||||| :||| :|||
765 .....ValGLUgluProser..... 769
435 ACAGGGCGCGGCTATCCCGCTCCCAAAGCGCGAGGATATATACAGCT 484
769 .....
485 ACGACATAAAGCGCTGGCCAAATATCCGCTCAGCTGACGACAAAC 534
||||| :||| :|||
770 .....ValProserProAsp..AlaP 776
535 CGCAG.....CACGGACACAGGCTTTCGACCGTTTCACAAATAC 575
||||| :||| :|||
776 roGLIIEthThrHisArgIleValGLUphArgAlaSerSerArgSer 792
576 CGGTAGTATGCTGACGCAAGAGTAGGGGACGATCAACGCGCCACCC 625
||| :||| :|||
793 ArgSerThrSerSerSerArgSerArgLysLysLysLysLysLys 809
626 GATA.....CAGCCCGAGCTGGACAGCA 648
||| :||| :|||
809 sValAlaArgGLUHisGLNArgThrArgSerSerThrArgSerGLys 826
649 TCGGCAATGCCGCGAAGCTTCAACGGGACGCTGACATATCGTCAAAA 698
||| :||| :|||
826 rGAspArgThrSerArgSerValSerProPheThrGLUHisThrLys 842
699 CATTCGCGCGGCGGCGAGAAATGTGCGGACGAGCGATGCGCAGG 748
||||| :||| :|||
843 ArgHisArgAlaLysThrLysSerArgArgSerSerAspArgAlaSe 859
749 GTATAAGCCAGAGCTCAAAACATTGCTTATGACAGGCTTGGGCTGCTT 798
||| :||| :|||
859 rSerGLNAspArgAlaLysArg..... 866

```

```

799 TCCACCGAAACAGATGGCGCGCATCAACGATTTGCGAGATATGCGCA 848
||||| :||| :|||
867 ..ArgLysAspArgAspArgGLUHisArgArg...GLYProThrPLy... 880
849 ACTCAAGACTGTGCCCGACGACCATCCGCGATTGGGCGCAACCAAC 898
||| :||| :|||
881 HisGLYArgCysThrArgLysSerArgSerArgSerArgProGLYSe 897
899 CCAATGCCCGCACAGCATAGAACGGCTGACGATATATCT.....TCG 942
||| :||| :|||
897 rSerSerCysGLUArgHisGLISerArgArgArgLysArgAlaHisSer 914
914 LysArgSerArgArgLysArgAspGLYSerProHisSerSerLeuGLUArg 930
993 GGGCGCATCAGCGGCGACATCCCTGTCAAG..... 1020
||| :||| :|||
931 AspArgArgHis..LysHisArgGLUArgSerArgArgLysArgMetAspLysL 947
1021 .....CGGTCCGAGATG..... 1033
947 YsGLUserMetThrArgSerArgArgLysArg..ArgArgThrArgSerArgSe 963
1034 .....GCGAGATCGCATTCGCGAAAGGAAATC 1061
||| :||| :|||
963 rProserValGLUHisArgThrArgArgProHisSerArgLysHisP 980
1062 CGCCGTC.....GCGACAAATTTGCGCGATGGCGGATCGGCA 1099
||| :||| :|||
980 roHisSerProGLUArgLysGLYAlaValArgGLUValSerProAlaPro 996
1100 .....AATACCCGTC... 1111
997 AlaProGLNGLYgluProArgGLNAspArgLysArgHisSerThrLysProBr 1013
1112 .....CTTACCATTCGCCGAATATCCGTTCAAACTTGAGACA 1148
||||| :||| :|||
1013 oValSerGLUValSerValLeuProGLUVal...ValSerValLeuProG 1029
1149 GCGTTACGGCAAGAAAGAACATCATCTCTCAACCGTCCGCGCTCAACG 1198
||| :||| :|||
1029 LuValValValAlaAspLeuAspProProGLU...ValProProValLeu 1045
1199 GAAAGATGTGAACCTGGCAACAAAGCCACCGAGACCAAAATGCGG 1248
||||| :||| :|||
1045 laGLUserValSerCys.....ValPro 1052
1249 TTTGAC.....GTTAAAGGTTTCGGA 1271
||| :||| :|||
1053 GLUAspLeuAspTyrGLYAspSerValGLUAlaGLYHisValPheGLUAs 1069
1272 TTTTGAAGAAAGAC.....GTAAATACGATACGAGAAATTAATACCG 1312
||||| :||| :|||
1069 pHisSerAsnGLUAlaIlePheIleGLINLeuAspAspMetSerSerProp 1086
1313 CTGTACCACA.....GTGANT 1329
||| :||| :|||
1086 roSerProGLUserThrAspSerSerProGLUArgAspPheLeuProAsn 1102
1330 CTTATAGTAAACCGGCTTAACTTAAAGGTTCT..... 1365
||||| :||| :|||
1103 ProIleLeuProProAlaSerLeuProGLNAsnSerThrLeuProValTh 1119
1366 .....GTGCGATCGGCTCATCTTGGCTATACCTCCAGAAATTC 1405
||| :||| :|||
1119 rGLNArgGLUValLeuProIleHisSerGLUAspIleSerLysProAlaP 1136
1406 AATACGCAAAATTACCA.....AGCAAGGTAGAAATC 1437
||| :||| :|||
1136 roGLNProLeuAlaProSerAspGLNcysLeuLeuArgGLNAspThrVal 1152
1438 AGATATATCCACCTAAATAATTACTCTCTTCAGACACCGCTACCAAAAG 1487

```

```

1153 GluThrThrAlaThrThrLeuSerThrProGluVal...LeuProMetG1 1168
1488 ACCTAATATATGATATTGGATAAATTGGTATATGATGACTAAA... 1533
1166 ylyaspserProLeuLeuSerGlyArgGlyCysGluValValArgProL 1185
1534 .....GATCCA.....TCAGAAGCTAAAGCTGCA 1557
1185 yAspAlaValAlaProAlaProLeuLeuArgSerArgThrLeuVallys 1201
1558 GAATTGGAATGGAGATTGCA..... 1578
1202 ArgValAlThrTrpAsnLeuGlnGluAlaGluAlaSerThrProAlaLeuAs 1218
1579 .....TTGTCTAAACAGAGAGAGAGAGACTTGA.. 1608
1218 PArgAspProArgThrProLeuGlnArgProGlnArgProGlnGluGlyA 1235
1609 ..TGGGCTAGTAGGAT 1623
1235 sPTpAspAlaGluAsp 1240

```

seq_name: sp_virus:056854

```

seq_documentation_block:
ID 056854 PRELIMINARY; PRT; 856 AA.
AC 056854;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DIC-2001 (TREMBlrel. 19, Last annotation update)
DE GP340.
GN BLF1.
OS Human herpesvirus 4 (Epstein-Barr Virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
ON NCBI_TaxID=10376;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M1;
RC MEDLINE=97082049; PubMed=8923292;
RA Mackett M., Cox C., Pepper S.D.V., Lees J.F., Beverley B.A.,
RA Wedderburn N., Arrand J.R.;
RT "Immunization of common marmosets with vaccinia virus expressing
RT Epstein-Barr virus (EBV) gp340 and challenge with EBV."
RL J. Med. Virol. 50:263-271(1996).
DR EMBL: X99106; CAA67558.1;
SQ SEQUENCE 856 AA; 89634 MW; E34E581AF1071CDD CRC64;

```

alignment_scores:

Quality:	139.50	Length:	599
Ratio:	0.589	Gaps:	26
Percent Similarity:	39.566	Percent Identity:	21.202

alignment_block:

US-09-303-518D-465 x 056854 ..

Align seg 1/1 to: 056854 from: 1 to: 856

```

110 TCAGCGTCGACGATTTCGAACCCGAGCGGAATAATCAACCTATTG..... 154
318 AsnThrThrAspIleThrValGlyAspAsnAlaThrTrpSerValPr 334
155 .....GCAGCAGGGGGAACTTCCGAGCGCAGCGCATATACG 154
334 oMetValThrSerGluAspAlaAsnSerProAsnValThrValThrAlaP 351
195 ATTCG.....GAAACATPACAAAGCCATCACTGGGCA 226
351 heTrpAlaTrpProAsnAsnThrGluThrAspPheLysCysLysTrpThr 367
227 ACCTGTCATCCAGCAGCGGCATTAAGCAATATCGGCTCATATTGTC 276

```

```

368 LeuThrSerGlyThrProSerGlyCysGluAsnIleSerGlyAlaPheAl 384
277 CGCT.....TTCCGATCACGGGC..... 295
384 aserAsnArgThrPheAspIleThrValSerGlyLeuGlyThrAlaProL 401
296 .....ACGAAGTCATTCCTCCCTTCGACACACCATGCTCAC 331
401 yStrLeuIleIleThrArgThrAlaThrAsnAlaThrThrThrHnHs 417
332 ATTCGATTCGTGAGAGCGGAGTAGTCCCTGACGATTCACCTTTAC 381
418 LysValIlePheSerIysAlaProGluSerThrThrThrSerProThrle 434
382 CGCATTCATTGGG...ACGGATACGACACCATCCCGCGAGCGATATGA 428
434 uAsnThrThrGlyPheAlaAspProAsnThrThrThrGlyLeuProSerS 451
429 CGGGCCACAGGGCGGCGCTATCCGCTCCCAAGCGGAGGATATAT 478
451 erThrHnHsValProThrAsnLeuThrAlaProAlaSerThrGly..... 465
479 ACAGCTACGACATAAAGCGCTTGCCCAAAATATCCGCTCA.....AC 522
466 .....ProThrValSerThrAlaAspValTh 474
523 CTGACGACACACCGCAGCAGCAGCGCTTGTGACCGTTTCCACAA 572
474 rSerProThrProAlaGlyThrThrSerGlyAlaSerProValThrProS 491
573 TACCGGTAGTATGCTGACGCAAGAGTAGCGGCGATTCACAAAGCGCCA 622
491 erProSer.....ProArgAspAsnGlyThrGluSerIysAlaPro 504
623 CCGGATACACCCCGAGCTGG..... 643
505 AspMetThrSerProThrSerAlaValThrThrProThrProAsnGlyTh 521
644 .....ACAGATCGGGCGAATCCGCCGA 665
521 rSerProThrProAlaMetThrThrProThrProAsnAlaThrSerProt 538
666 ACCTTTCACAGCGCAGCATGCAATATCGTCAAAACA..... 700
538 hreulGlyStrThrSerProThrSerAlaValThrThrProThrProAsn 554
701 .....TCATC 705
555 AlaThrSerProThrProAlaValThrThrProThrProAsnAlaThrSe 571
706 GGGCGGCGAGGAATTTGTGCGCGCAGCGCATGCCGTGACGATATAG 755
571 rProThrValGlyGluThrSerProGlnAlaAsn.....A 583
756 CGAAGGCTCAACATG.....CTGTATGCAAGCGCTTGGCTGCTT 799
583 lathrAsnHsThrLeuGlyGlyThrSerProThrProValValThrSer 599
800 CCACCGAAAACAAGATGGCGGCATCAACGATTTGGCAATATGGCGCA 849
600 ProProLysAsnAlaThrSerAspValThrThrGlyGlnHsAsnArgTh 616
850 CTCA.....AGACTATGCCGACGACGACCATCCCGGATGGCGAGT 890
616 rSerSerThrThrSerMetSerLeuArgProSerSerIleProGluT 633
891 CCAAAACCCCAATGCCGACACAGCATAGAACCGCTCAGCAATATCTTTA 940
633 hTrThrSerHsMetPro.....LeuLeu 640
941 CGGCAGTATCCCGTCAAGAGGATTTGAGCTTTGGGGAATAATACGC 990

```

```

641 ThrsrAlahsProthrGlyGlyGluAsnIlethrGlnValThrProal 657
991 TTGGGCGGCATCAGGCACATCTGTCAAGCGGTGAGATGGGCGAGAT 1040
657 aserIleSerThrIshIshValSerThrSer..... 667
1041 CGCATTCGCCAAGGAATCCGCCGTACAGCAATTTCCGATCCGG 1090
667 ..... 667
1091 CATACGCCAATATACCCCTCCCTTACCATTCGCCAATATCCGT..... 1134
668 ..SerProAlaProArgProGlyThrThrSerGlnAlaSerGlyProgl 683
1135 ....TCAAACTGGAGCAGCGTTACGGCAAGAAACATACCTCCCTCAA 1180
683 yAsnSerSerThrSerThrLysProGlyGluValAsnValThrLysGlyT 700
1181 CCGTCCCGCGGTCAACGGAAGAAATGTGAACACTGCCAACAACGCCAC 1230
700 hr...ProProLysAsnAlaThrSer..... 707
1231 CCGAAGACCAAAAGTCCCGTTGACGGTAAGGTTTCCGAATTTGAAA 1280
708 ProGlnAla.....ProSerGlyGlnly 715
1281 AGAGCTAAATATCGATACGAGATTAATACCGTGTACCCACAGTATC 1330
715 s.....ThrAlaValProThrValThrS 723
1331 CTATAGATGAACCGCTTTTAACTCAAGGTTCTGTCGATCCGCTCAT 1380
723 erThrGlyGly.....LysAlaAsnSerThrThrGlyGlyLysHIS 736
1381 TCTGTGCTAATACTCCAGAAATT..... 1404
737 ThrThrGlyHISglYAlaArgThrSerThrGluProThrThrAspTyrG 753
1405 .....CAATACGCAAAATTAACCAAGCAGTGAATCAGATATATACC 1447
753 yAspAspSerThrThrProArgProHISgTyAsnAlaThrThrLysLeuP 770
1448 CACCTAAATTAATCTCTCTTACGACCGCTACCAAAAGACCTAATAT 1497
770 roProSerThrSerSerLysLeuArgPro..... 779
1498 GGATTTTGGATTAATTTGGTAATGAATGACT...AAAGGCCATCAAG 1544
780 .....ArgThrThrThrSerProProva 788
1545 AACTAAGGTCAAGAAATTTGAATGGATTTCAATGTCTAAACAGGAA 1594
788 lThrThrAlaGlnAla.....ThrValProValProProThrSerG 802
1595 GAGAGCAA.....CTGGATGGGCTAGT 1617
802 lnrProArgPheSerAsnLeuSerMetLeuValLeuGlnTrpAlaSer 817
seq_name: sp_invertebrate:Q17921
seq_documentation_block:
ID Q17921 PRELIMINARY; PRT; 769 AA.
AC Q17921;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEtical 81.0 KDA PROTEIN.
GN C12D12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiodea;
OC Rhabditiidae; Peioderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; Pubmed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Nhan M., Hawkins J.;
RT "The sequence of C. elegans cosmid C12D12.";
RL submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U51998; AAA96080.2;
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KM Hypothetical protein.
SQ SEQUENCE 769 AA; 80970 MW; F922FA6F5FE3269 CRC64;

```

```

alignment_scores:
  Quality: 136.00      Length: 304
  Ratio: 1.062         Gaps: 9
  Percent Similarity: 42.105   Percent Identity: 24.013

```

```

alignment_block:
US-09-303-518D-465 x Q17921 ..

```

```

Align seg 1/1 to: Q17921 from: 1 to: 769

```

```

68 AGCCCTCACATTTGGCAACGATCTTTATCCGCGAGGTCTCGACCGT 117
   |||||
508 ThrProThrThrValGlyThrMetThrProSerThrGlyThrThrVal 524
   |||||
118 CAGCATTTGGAACCCGCGAGGAATACCACTTATCGCGAGGAGGGA 167
   |||||
524 lProValProThrThrThrGlySerProThrThrGln..... 536
168 ACTTCCGAGCGCGCGGTCAATCGATTTGGGAACATACAAAGCATC 217
536 .....
218 AGTTGGCAACCTGTTCATCCAGCAGCGCCATTAAAGAAATATCCGC 267
   |||||
537 .....ThrThrAlaProValThrLysProThrValProSerSerTh 550
268 TACATGTCCGCTTTCCGATCGAGCGCAAGATCCATCCCTTCCA 317
   |||||
550 rThrGlnThrAlaProProValThrThrProThrSerGlnProProVal 567
318 CAACCATGCTCACATTCGATTCGATGAGCGGTAGCCGCTGAGC 367
   |||||
567 hrThrThrSerLeuLeuThrThrLeuThrThrProThrAlaProValThr 583
368 GATTCAGCTTTACCGCATCCATTTGGAGCGATACAGACCATCCCGCC 417
   |||||
584 ThrThrValValProSerSer.....AlaThrValProThrThrProPr 598
418 GAGCGCTATGAGCGCCACAGCGCGGCTATCCGCTCCCAAGCGC 467
   |||||
598 ofThrThrValThr.....ValAlaAlaThrThrThrSerLysAlaP 612
468 GAGGATATATACAGCT.....ACGACATAAAGCGCTTG 502
   |||||
612 roValValThrThrSerProThrLeuAlaProThrSerProThrLysLeu 628
503 CCCAAATATCCGCTTCACCTGACCGACAAACGACGACCGGACAGG 552

```

```
629 ProThrSerProProSerThrValGlyThrSerProThrAlaProAlaAs 645
    ||| ||||| |||::: :::
553 CTTGTCGACCGTTTCCACATACCGGATGCTGACCGAAGAGTAGG 602
    :||:::|||| :||| |||||
645 nLeuThrThrProThrThrAlaProValasn.....P 656
603 CGACGGATTCAAGCGCCACCC.....GATACGCCCCG 637
    |||:::||||: ||| |||:::||||
656 roThrSerSerThrThrAlaProThrAlaProValasnProThrSerPro 672
    :||: ||: |||: |||: |||:
638 AGCTGGACAGATGGGCAATGCCCGCAAGCTTTCACAGGCGACTGCAGAT 687
    :||: ||: |||: |||: |||:
673 ThrThrAlaProThrValProProValThrThrThrProThrThrTh 689
688 ATCGTCAAAAACATCATCGCGCGCAGAGAAATGTCGGCGCAGCGA 737
    :||| |||:::||||:
689 rThrSerThrThrThrThrThrThrThrThrThrThrGlnThrT 706
738 TGCCGTGCAGGGTATAGCGAAGGCTCAAAACATGCTGTATGACGCGCT 787
    ||| |||
706 hrPro.....ThrThrProValValThrThr 714
788 TGGGTCTGCTTCCACCGAAACAGATGGCGCGCATCAACGATTGGCA 837
    :||: ||| |||||
715 ProSerThrIleThrProThrThrArg..... 723
838 GATATGGCGCAACTCAAGACTATGCGCGCAGCCATCGCGATTGGGC 887
    |||:::||||: |||
724 .....ProValThrGlnProAlaSerGlnProAlaThrGlnP 736
888 AGTCCAAAACCCCA.....ATGCCGCAAGGCAATAGAGCGCGTCA 928
    :||: ||||| |||: |||
736 roAlaThrThrProThrThrThrIleProThrThrThrGlnPheSerSer 752
929 GCAATATCTTTA 940
    |||:::||||:
753 AlaValSerVal 756
```

